Culture Scientifique-Biologie
« Penser/Traduire »

Sommaire

Programme et modalités de contrôle des connaissances **2**

Cours 1 : Penser/Traduire, définitions **3**

Cours 2 : La représentation « objective » du monde **5**

Cours 3 : Comment décrire le monde ? **7**

I. La méthode expérimentale 7

II. Matière à penser : état de l'art dans un domaine scientifique 7

1) La collecte de la documentation 8

2) Le cadre conceptuel 8

Cours 4 : Le langage de l'ADN. Conservation, transmission et traduction de l'information génétique **14**

I. ADN, une molécule universelle et variable 14

II. Reproduction conforme de la cellule et réplication de l'ADN 14

III. Variabilité génétique et mutation de l'ADN 15

IV. L'expression du patrimoine génétique 16

V. Epigénétique et méthylation 18

Cours 5 : Les bases de l'unité et de la diversité du vivant, concept d'espèce et arbres phylogénétiques **19**

I. Bref aperçu de l'histoire des concepts en évolution. La nomenclature du vivant. Théories pour l'origine et l'évolution des espèces. 19

II. Notions de phylogénie et arbres phylogénétiques du vivant. 25

a. Utilisation de critères morpho-anatomiques. 25
b. Utilisation de critères de développement 26
c. Utilisation de critères moléculaires. 26

III. Que nous indique un arbre phylogénétique ? 27

Cours 6 : Les bases de l'unité et de la diversité du vivant, concept d'espèce et arbres phylogénétiques **29**

I. Diversité génétique, sélection et reproduction. 29

a. Cycle chromosomique, cycle de vie, reproduction sexuée et méiose. 29
b. Sélection naturelle et reproduction sexuée. 32

II. Code barre moléculaire du vivant. 33

Cours 7 : Génétique et éthique **37**

I. Les maladies génétiques. 37

a. Les maladies monogéniques 37

b. Le cas du cancer du sein 39

II. Notions de génétique moléculaire des cancers. 40

a. Bases moléculaires de l'oncogenèse 40
b. Agrégation familiale de tumeurs fréquentes..... 40

III. Génétique et éthique. 41

a. Cadre de définition 41
b. L'éthique de la médecine..... 42
c. La génétique médicale..... 42
d. La procréatique et le diagnostic prénatal/pré-implantatoire..... 44
e. Le clonage et la modification génétique..... 45
f. Les expérimentations sur l'homme 47

Cours 8 : Les faux-semblants **47**

I. Evaluation statistique de la similarité et de la différence. 47

a. Métrologie 47
b. Variabilité technique et biologique 48
c. Comparaison de moyennes 50
d. Intervalle de confiance 52

II. Réflexions autour de la corrélation et du lien de causalité. 52

a. Corrélation et régression 52
b. Notion de covariance 54
c. Notion de coefficient de corrélation 54
d. Corrélation, cause, effet, multicausalité 55

III. Homologie et analogie. La convergence évolutive aux différents niveaux d'intégration du vivant. 56

a. Notion d'homologie 56
b. Notion d'analogie..... 56
c. Notion de convergence évolutive 56
d. Notion de réversion 57

Programme :

Cours 1 : Penser/Traduire : définitions.
Cours 2 : La représentation « objective » du monde.
Cours 3 : Comment décrire le monde ?

I. L'utilisation de la méthode expérimentale.

II. Matière à penser : état de l'art dans un domaine scientifique.

Cours 4 : Le langage de l'ADN : Conservation, transmission et traduction de l'information génétique
Cours 5 : La réinterprétation des données et le changement de cadre conceptuel. Les bases de l'unité et de la diversité du vivant, concept d'espèce et arbres phylogénétiques.
Cours 6 : Diversité génétique et code barre moléculaire.
Cours 7 : Génétique et éthique.
Cours 8 : Les faux semblants.

I. Evaluation statistique de la similarité et de la différence.

II. Réflexions autour de la corrélation et du lien de causalité.

III. Homologie et analogie. La convergence évolutive aux différents niveaux d'intégration du vivant.

Cours 9 : Expérimentation animale et médecine.

I. Les quatre étapes de la cruauté.

II. Spécisme, anthropocentrisme, biocentrisme et écocentrisme

III. Modèles animaux et expérimentations.

Modalités de contrôle des connaissances : 2 contrôles en continu (12 octobre et 15 décembre)

Cours 1 : Penser/Traduire, définitions

CONCEPT

Idée générale et abstraite que se fait l'esprit humain d'un objet de pensée concret ou abstrait, et qui lui permet de rattacher à ce même objet les diverses perceptions qu'il en a, et d'en organiser les connaissances.

Un concept est une idée générale qui permet de regrouper toute la diversité d'un objet autour d'une même unité de pensée, pour mieux la manipuler. Prenons par exemple le concept de l'amour : c'est une idée générale, abstraite qui ne peut se définir en dehors de toutes les expériences particulières et subjectives de ce concept. Autrement dit, même si chacun a sa « vision » de l'amour, tout le monde l'expérimente et connait tous les éléments qui permettent de penser l'amour.

CADRE CONCEPTUEL

Un cadre conceptuel est un outil d'analyse comptant plusieurs variations et contextes. Il est utilisé pour faire des distinctions conceptuelles et organiser des idées. Les cadres conceptuels forts saisissent quelque chose de réel et le font d'une manière facile à retenir et à appliquer. Le terme « cadre conceptuel » s'emploie à la fois selon l'échelle (grandes et petites théoriques) et les contextes (sciences sociales, marketing, science appliquée, art, etc.). Sa définition et son application explicites peuvent donc varier.

Les cadres conceptuels sont particulièrement utiles comme organisation de dispositifs dans la recherche empirique. Un certain nombre de chercheurs appliquent la notion de cadre conceptuel à la recherche empirique déductive.

Il y a plusieurs types de cadres conceptuels et ils se conforment tous à un but de recherche : hypothèse de travail, exploration ou recherche exploratoire (une recherche à caractère d'exploration), catégorie descriptive (la classification des plantes de Carl Von Linné par exemple), description ou recherche descriptive (recherche dont le caractère ou le but est de décrire), type d'idée pratique, analyse, modèles de recherche opérationnelle, prise de décision, hypothèse formelle, explication et prédiction.

Un cadre conceptuel est l'outil que nous employons pour mettre en ordre, organiser, penser le matériau brut de notre recherche (données, observations empiriques, références…). C'est le terreau de la recherche en ce sens où il va permettre la création, l'articulation et la bonne croissance de notre recherche. Le ou les cadres conceptuels veillent à l'avancée de la recherche pour mettre en exergue et résoudre les difficultés rencontrées, paradoxes, apories ou toute situation à interprétation.

VISION DU MONDE

La vision du monde désigne, en philosophie, en psychologie sociale, l'ensemble des représentations métaphysiques, physiques et sociales qui orientent l'action des êtres humains.

La vision du monde est l'ensemble des racines, des noyaux durs de notre connaissances du monde. Ce sont toutes les croyances fondamentales, présuppositions, et axiomes que l'on a sur le monde et notre existence. La vision du monde est une expérience subjective, unique et dirige nos actions. C'est parce que je me fais telle idée sur le monde que je vais agir de sorte. Par exemple, si je pense que dans le monde, les forts écrasent les faibles, je vais faire en sorte d'écraser autrui et tout mes potentiels rivaux pour pouvoir faire partie de la catégorie des forts, catégorie la plus apte à la survie car elle ne subit aucune action négative selon ma vision du monde.

PENSER	TRADUIRE
Du latin *pensare* (peser, soupeser) • Considérer quelque chose comme vrai (ou non), comme probable (ou non). • Considérer quelqu'un, quelque chose comme étant tels (*Je le pensais diplomate*.). • Avoir telle ou telle opinion sur tel sujet. • Avoir la conviction de quelque chose, le croire. • Avoir l'intention de, projeter. • Faillir, manquer (*Elle pensa s'évanouir*). • Se dire intérieurement. • Comprendre, déduire. • Concevoir quelque chose dans ses moindres détails avant de le réaliser. Synonymes : admettre, croire, estimer, imaginer, juger, conjecturer, espérer, se figurer, supposer, élaborer, étudier, examiner, mûrir, ordonnancer, pénétrer, régler	Du latin *traducere*, « faire passer d'un point à un autre » • Transposer un discours, un texte, l'exprimer dans une langue différente • Exprimer un sentiment, une pensée, les rendre sensibles Synonymes : Transposer, manifester, montrer, révéler, trahir, convoquer (traduire quelqu'un en justice).

➔ Comment ces termes peuvent s'articuler dans le domaine des sciences ?

Penser en sciences, c'est considérer quelque chose comme vrai (ou non) ou comme étant tels, avoir la conviction de, le croire ; au travers d'une compréhension, d'une déduction, d'une conception dans ses moindres détails.

Traduire en sciences, c'est exprimer, rendre intelligible, interpréter, modéliser et idéaliser le monde qui nous entoure grâce à l'activité de notre cerveau afin de se le représenter dans ses moindres détails et de partager avec les autres son interprétation.

Il y a donc un passage entre individus/cerveaux et donc une trahison possible !

Cours 2 : La représentation « objective » du monde

➔Peut-on avoir une représentation « objective » du monde ?

Document 1 : Le corbeau et l'interprétation différentielle (vidéo)

De prime abord, le corbeau semble éloigné de nous, semble avoir moins d'importance. Cependant il possède une capacité à mémoriser et décortiquer les différentes étapes d'un problème pour le résoudre. Faculté que l'on pensait jusque là uniquement possible chez le cerveau humain. Moins un être vivant est capable de faire démonstration d' « intelligence » (ou autrement dit, de mimer nos capacités d'actions et de réflexions), moins nous allons lui accorder une valeur intrinsèque dans le jeu complexe de la vie. L'expérience du corbeau montre qu'il est capable d'utiliser des outils pour saisir des objets, qui eux-mêmes vont servir à actionner un mécanisme complexe pour faire parvenir la nourriture. C'est un outilleur fin insoupçonné.

Document 2 : Citation de Nietzsche sur le rêve

« C'est en effet grâce à deux états que l'homme conquiert la sensation d'exister, le rêve et l'ivresse. La belle apparence du monde des rêves, au sein duquel tout homme est pleinement artiste, est la mère de tout art plastique… » Friedrich Nietzsche, *La vision dionysiaque du monde*.

Pour Nietzsche, le rêve et l'ivresse permettent de transcender la réalité pour véritablement exister. Ces deux modalités d'être permettent de s'affranchir d'une réalité ascèse et aride pour s'accomplir en tant qu'artiste. En effet, l'art permet à l'homme de s'accomplir lui-même, l'art donne à l'homme la capacité de devenir un « surhomme » c'est-à-dire celui qui jouit de son être de manière pleine et entière. Une représentation « objective » du monde n'est pas ce qui importe a priori. C'est paradoxalement par le rêve et l'ivresse que l'on peut accomplir de la manière la plus complète et objective, son humanité.

Document 3 : L'addiction ou la vérité modifiée. *Nuggets*, court métrage d'animation d'Andreas Hykade (2014)

Synopsis du site « Sens Critique » : « Un petit oiseau, une sorte de kiwi ou de pingouin trouve une jolie pépite dorée. En l'avalant, il s'envole dans les airs avec délectation. Mais tandis que notre oiseau continue sa quête de pépites, à la poursuite de ce premier état de volupté, son univers s'assombrit peu à peu. Au fur et à mesure, il s'envole de moins en moins haut et sa chute devient de plus en plus dure... »

Le court-métrage parle du processus de l'addiction et de son problème : de la curiosité à l'extase pour finir par le besoin destructeur. Si les « pépites » permettent au protagoniste d'accéder à un monde pus agréable, meilleur et supérieur, leur effet devient de moins en moins fort. En s'enfonçant dans l'accoutumance, le protagoniste modifie sa réalité. Pour lui, le monde réel devient un cauchemar et c'est inlassablement qu'il va revenir vers les « pépites » pour retrouver un monde meilleur qu'il va prendre pour sa réalité.

Document 4 : Interprétation et falsification (Journal télévisé de France 3, 19h, 15 Décembre 2018, accusation de falsification d'image)

Aux premiers abords, sans connaître le contexte de l'image, on pense à une manifestation de soutient sans la pancarte complète. En réalité, l'image a été modifiée et le « dégage » a été enlevé de la pancarte « Macron dégage ».

Il y a donc une falsification de la pensée par falsification du réel. La trahison n'a pas lieu dans l'interprétation des données mais par les données elles-mêmes qui sont incomplètes ou falsifiées. A part des données du réel ou fabriquées, notre cerveau va penser et traduire des données pour diriger nos actions.

Document 5 et 6: Le développement des animaux et les générations successives (Théorie de la préformation, 1694, *Essai de diotrique*, Nicolas Hartsoeker et étude de l'embryologie (formation de l'embryon de poulet) par Marcello Malpighi (1673)).

La théorie de la préformation est une théorie formulée au XVIIIème siècle pour expliquer le développement embryonnaire par le déploiement de structures déjà existantes dans l'œuf, à la manière de poupées russes. Deux modèles s'affrontent : l'ovisme (l'embryon est préformé dans la femelle) et l'animalculisme (l'embryon est préformé dans le spermatozoïde). Hartsoeker se situe dans ce dernier modèle avec le concept d'« homoncule ». Il imagine un fœtus entier qui pouvait se loger dans la tête du spermatozoïde, comme une réplique miniature.

Malpighi explique la formation de l'embryon de poulet comme le résultat de plusieurs couches cellulaires qui vont se développer et former l'embryon. Ces couches sont présentes dans l'œuf et vont juste continuer à se former petit à petit. Tandis que Hartsoeker était partisan de la théorie de la préformation, Malpighi était plus proche de sa théorie rivale, l'épigénèse. Cette théorie du XVIIIème siècle stipule que l'embryon se développe petit à petit de manière complexe en rapport avec son environnement.

Que retenir de ces deux théories ? Qu'il est difficile d'avoir une représentation objective du monde. Nous pouvons uniquement s'approcher au mieux de la vérité des choses par le biais de comparaisons, métaphores ou liens forcés selon nos capacités d'observation du réel et nos connaissances actuelles.

Document 7 et 8 : Expérience réelle *versus* simulation numérique (simulation de larve de poisson zèbre)

De façon générale, la simulation numérique est un outil qui permet d'étudier les résultats d'une action sur un élément sans réaliser l'expérience sur l'élément réel. Cette solution alternative permet de lever divers problèmes de la réalisation de l'expérience réelle et de prévoir les résultats. Quelle que soit la qualité de la simulation, elle ne remplace pas totalement l'expérimentation. Une simulation numérique est utile pour la prédiction et l'évaluation, à la condition qu'il existe une reproductibilité acceptable entre le modèle numérique et le modèle réel.

La simulation numérique reproduit ce que l'on est capable d'observer, l'observation permet d'optimiser la simulation numérique. La simulation numérique nécessite la mise en équation du réel. Cette mise en équation devrait nous permettre d'utiliser le modèle numérique afin d'obtenir des paramètres chiffrés supplémentaires de ceux obtenus avec l'expérience réelle.

Conclusion générale : Peut-on plaider une représentation objective du monde ? Au regard des nombreux paramètres qui faussent notre pensée, il semble a priori que non : rêves, ivresse, jugements de valeurs, préjugés, connaissances limités, biais cognitifs, spécisme, réalité modifiée, interprétation, falsification…Cependant, il semble que la mise en équation du réel puisse permettre d'approcher au maximum une représentation « objective » du monde.

Cours 3 : Comment décrire le monde ?

I. La méthode expérimentale

La méthode expérimentale s'effectue dans le cadre d'une recherche empirique déductive. La recherche empirique peut être définie comme « la recherche basée sur l'expérimentation ou l'observation (évidence) » (Claude Bernard, 1865). Ce type de recherche est mené avec l'objectif de tester une hypothèse.

Nous pouvons schématiser la méthode expérimentale ainsi :

(Modèle théorique ➔ Prédiction)

⬇

Observations➔Question➔Hypothèse➔…➔ Expérience(s) ➔Résultat(s) ➔Confrontation à l'hypothèse➔Conclusion(s).

Un raisonnement **déductif** part d'une idée générale, d'un principe, d'une loi pour en tirer une conséquence particulière. Pour discuter le raisonnement, on peut analyser la valeur de la loi générale, repérer si le fait particulier entre bien dans le domaine de la loi générale.

Un raisonnement **inductif** part d'un ou de plusieurs faits particuliers pour en tirer un principe, une loi, une idée générale. Ce raisonnement est inverse au précédent. Pour discuter le raisonnement, on analyse la pertinence de l'extension du fait particulier à un ensemble plus vaste.

II. Matière à penser : état de l'art dans un domaine scientifique= l'art de l'exhaustivité : encyclopédies bibliographiques et biologiques et outils informatiques.

Aucune recherche, aussi originale soit-elle, n'est isolée par rapport à un **signifiant**. Elle se rattache toujours à un **contexte de sens**. Le cadre conceptuel a pour but de cerner cet ensemble idéal (théories, lois, concepts, données scientifiques).

Que la recherche soit entreprise pour apporter un éclairage nouveau sur une question, pour construire un corpus explicatif, pour approfondir une analyse, dans tous les cas, elle fait **progresser la connaissance** sous un aspect ou dans un domaine inédit.

Avant de découvrir et d'affirmer qu'il s'agit bien d' « inédit », le chercher doit faire le point sur le sujet qu'il étudie, **explorer** le champ (intellectuel) où il veut travailler, **considérer** les constructions théoriques déjà énoncées dans le domaine. Sa démarche débute par une recherche documentaire très vaste puis par une analyse de cette documentation à partir de laquelle il bâtit un **cadre conceptuel**.

1) La collecte de la documentation
    a. Recenser les sources d'informations bibliographiques.

Elles doivent être consultées systématiquement en partant des plus récentes. Le chercheur relève les auteurs, les documents qui l'intéressent. Ceux qui retiennent son attention sont fichés. Il peut consulter les publications, livres, périodiques, textes législatifs, publications administratives, thèses, mémoires, traductions, statistiques, annales, documents professionnels, photographies, films…

b. Cerner l'information

Cette opération est particulièrement délicate car il convient pour le chercheur de se procurer une bibliographie à la talle de sa problématique. Mais où cela s'arrête-t-il ? Il peut s'appuyer sur son hypothèse de recherche pour en extraire des **mots-clés** et procéder à une sorte de **thésaurus**.

« Les thésaurus sont des langages contrôlés constitués de « Descripteurs » mots ou expressions, pouvant se combiner entre eux »

C'est un outil documentaire où un mot clé renvoie à un certain nombre **d'autres termes**. Cette opération permet également de **préciser le sens** de chaque signifiant et **d'éviter la polysémie** (caractère d'un signe qui possède plusieurs contenus, plusieurs sens…).

Cerner l'information revient à la **sélectionner**, à rechercher ce qui est inédit, original, à laisser tomber les travaux qui ne sont que paraphrase, à écarter les pistes secondaires, mais aussi à ne rien oublier de ce qui a déjà été écrit sur le sujet.

c. Exploiter l'information

Tous les documents consultés ne sont pas mémorisés d'emblée. Ceux qui retiennent l'intérêt du chercheur sont **condensés**, c'est-à-dire qu'il extrait l'information utile contenu dans le document et en fait un **résumé** (abstract) qui l'aidera pour la mémorisation. Les abstracts sont **classés par thèmes** ou **mots clés**. Chacun a une méthode qui lui est propre pour classer cette information.

Quand la bibliographie a été fructueuse, le chercheur se trouve en face d'une importante quantité de matériaux. Il peut être submergé par le nombre et la diversité des documents. **Un tri s'impose** pour l'exploitation. Les appréciations et critiques, faites lors de la lecture, permettent de classer la bibliographie **par la valeur** que le chercheur lui donne : essentiel, très important, à ne pas négliger…, à mettre si je n'ai rien trouvé d'autre etc…

Cette méthode implique au préalable, une **« lecture intelligente »**, c'est-à-dire que les axes définis dans servent d'objectif de lecture.

d. Analyser l'information

Cette phase de la documentation prépare le **cadre conceptuel** de la recherche. Le chercheur est amené à se poser des **questions** : ai-je acquis des connaissances, qui orientent différemment la recherche ? Ai-je bien apporté un éclairage sur tous les aspects de la recherche ?

Pour trouver des éléments de réponse, il **analyse** sa documentation, fait des **liens** et des regroupements, met en **évidence** les contradictions, les courants de pensées, d'idées opposées, procède une critique des textes, etc… Cela est possible quand il a rassemblé suffisamment de matériaux. Le fait même de les collecter (les textes) en un ensemble qui lui est inédit, fait surgir un champ sur lequel vous pourrez réfléchir, argumenter, attester ou contester.

Ce temps d'analyse sert également à retenir les citations d'auteurs qui renforcent les idées que le chercheur soutient. Il n'est pas question de faire une compilation de tous les textes lus, mais de faire **émerger**, en les regroupant, les concepts, théoriques qui vont **étayer** le cadre conceptuel.

2) Le cadre conceptuel

On appelle**, cadre conceptuel**, l'ensemble des connaissances, théoriques, qui ont un apport quelconque avec le sujet de la recherche.

Ces éléments vont servir de points de repères et devenir cadre théorique ou encore ils peuvent être un idéal à atteindre et devenir cadre philosophique.

Dans les premières étapes, de cadre conceptuel oriente, enrichit la problématique. Il féconde les hypothèses pour leur donner **toute l'envergure de leur signification**, mais rien que leur signification. Il détermine **l'idée ou les idées directrices**. Il est le lieu de **l'articulation** de sens contenu dans les variables. Il ne procure pas d'explication aux phénomènes, mais il aide à leur compréhension en donnant accès à des références connues, lois et théories.

Dès le début de la recherche, le cadre conceptuel a une fonction **d'organisation**, dans le sens où il oriente la démarche de pensée et la logique des différentes étapes. Au cours du travail, le chercheur teste la validité conceptuelle de la recherche. Il va faire référence au cadre conceptuel qu'il a bâti. Pour mettre en évidence un nouveau concept, une nouvelle connaissance, il doit analyser les relations que celui-ci entretient avec les lois, théoriques et phénomènes connus. **On ne fait progresser le savoir qu'en construisant, à partir de connaissances existantes, une démarche d'analyse adéquate.**
Dans les dernières étapes, le cadre conceptuel est indispensable pour évaluer les résultats. On peut lui comparer les contenus significatifs recueillis par les outils de recherche et ayant déjà été soumis à une première étude. Plus le cadre conceptuel est adapté et complet, plus l'analyse est fine et subtile. Il n'y a pas de recherche valable sans cette double confrontation de l'hypothèse, puis de l'analyse au cadre conceptuel. Le travail sur le réel implique toujours une position théorique scientifique de départ.

a. Histoire simplifiée des banques de données biologiques et bibliographiques

Une banque de données biologique est la collecte, le stockage et l'organisation de la connaissance biologique. Les données sont associées avec les informations biologiques et sont utilisées à l'aide d'interfaces informatiques utilisateurs conviviales.

Deux types de banques de données biologiques :

- *Banques de données primaires :* elles contiennent les informations, par exemple, de séquences nucléotidiques ou protéiques, accompagnées des informations d'annotations à propos de leur fonction, de la bibliographie, et des références croisées avec d'autres banques de données, etc…
- *Banques de données secondaires :* elle résume les résultats des analyses des banques de données primaires de séquence des protéines. Le but de ces analyses est d'obtenir des propriétés communes pour des classes de séquences, qui à leur tour permettent la classification de séquences inconnues (annotation). De plus, toutes les autres banques de données qui sauvegarde de l'information biologique ou médicale *exempli gratia* (par exemple) banques de données bibliographiques, sont fréquemment classées comme banques de données secondaires.

Développement d'Internet, du WWW et de l'usage des ordinateurs :

- Création d'Internet : 1969. L'ARPA (Advanced Research Project Agency) crée sous la gouverne du Ministre de la Défense des Etats Unis, un réseau de communication entre ordinateurs, l'ARPANET. Sa propriété est qu'un défaut sur une partie du réseau ne remet pas en cause le fonctionnement de la totalité du réseau. Deux protocoles sont utilisés pour faire fonctionner le réseau :

Le TCP (Transmission Control Protocol) et l'IP (Internet Protocol). Internet ce sont les réseaux connectés entre eux par les protocoles de communication TCP/IP/

- Le premier ordinateur personnel est mis sur le marché par IBM en 1981.
- Le World Wide Web (WWW) se développe entre 1990 et 1991 avec Tim Berners-Lee, du CERN). Deux services dominent alors l'internet : WWW et l'e-mail (inventé en 1971 par Thomlinson). Ces services sont supportés par un protocole de communication, le FTP (File Transfer Protocol), développé initialement pour le transfert de données scientifiques.
  Les moteurs de recherche moderne (Internet Explorer, Mozilla Firefox utilisent les protocoles http mais également FTP et ont permis le développement de plateformes de communication.

De la génomique associée à la bio-informatique :

- Le premier algorithme de comparaison de séquences de protéines ou d'ADN a été publié par Needleman et Wunsch en **1970**.
- Le terme bio-informatique a été inventé en **1978** par Hogeweh et définie comme « studies of informatic processes in biotic systems » (les études des processus informatiques dans des systèmes biotiques (relatif aux êtres vivants et à leur action)).
- La PDB (Brookhaven Protein Data Bank) a été crée en **1971**. La PDB est une banque de données pour le stockage des données cristallographique des protéines.
- En **1977**, Sanger et Al développent une méthode de séquençage de l'ADN (enzymation chain termination procedure). La publication de la séquence complète du virus bactériophage FX174 en **1980** coïncide avec la première utilisation du premier ensemble de programme désignés pour l'analyse des séquences des protéines et des ADNs (IntelliGeneticsSuite).
- Smith et Watermann publient un autre algorithme de comparaison de séquences en **1981**.
- En **1982**, la banque de donnée publique GenBank est créée. 2000 séquences déposées en fin **1983**.
- Création en **1982** du Genetics Computer Group (University of Winconsin) avec la mise au point de packages pour la biologie moléculaire nommé Wisconsin Suite. IntelliGenetics Suite et Wisconsin Suite étaient utilisés par lignes de commande. Le descendant actuel du WS est disponible sous le nom Accelys GCG.
- En **1986**, Mullis et ses collègues publient le principe du PCR (Polymerase Chain Reaction), qui représente une étape clé en biologie moléculaire.
- Création en **1986** de la banque de données SWISS-PROT et du terme « génomique » qui décrit une discipline scientifique relative au séquençage et à la description de génomes entiers
- Création en **1988** du NCBI (National Center for Biotechnology Information) à partir duquel sont hébergés, encore aujourd'hui, les banques de données les plus importantes. La même année débute le « Human Genome Initiative » et est publié l'algorithme FASTA
- **1991**, G. Venter publie l'utilisation des « Expressed Sequence Tags » (ESTs)
- **1992**, Venter et Fraser fondent le TIGR (The Institute for Genomics Research)
- **1994**, publication d'un outil intégré d'analyse de séquence (GeneQuiz) qui a été utilisé en **1996** pour la première analyse automatique des plus de 6000 protéines déduites de l'analyse de la séquence complète du génome de la levure de boulanger *Saccharomyces cerevisiae*.
- **1996**, Bairoch et ses collègues annoncent le début de la banque de données « Prosite ».
- **1997**, LION Biosciences AG est fondé à Heidelberg en Allemagne et utilise comme base le package GeneQuiz pour son produit phare, le package d'analyse de séquence bioSCOUT puis développe le package Sequence-Retrieval System (SRS).

- **1997 et 1998** sont des années charnières en bio-informatique et en biologie moléculaire du fait que les séquences complètes du génome de deux espèces modèles importantes ont été publiés, *Caenorhabditis elegans* et *Saccharomyces cerevisiae.*
- En **2000**, la société Metalife AG en Allemagne est crée et développe une banque de donnée sémantique nommée BioExplorer. Il combine les banques de données primaires et secondaires, ce qui permet à l'utilisateur d'extraire uniquement les informations pertinentes et non redondantes. La séquence du génome de deux autres espèces modèles majeures sont mises à disposition : *Arabidopsis thaliana* et *Drosophila melanogaster*.
- **2001**, la première version brute du génome humain est publiée et est complétée en **2004**.
- Une séquence brute du génome de la souris *Mus musculus* (responsable de la malaria humaine) en **2002**, le moustique *Anopheles gambiae,* le rat *Rattus norvegivus* (**2004**) et le chimpanze *Pan troglodytes* (**2005**).
- En **2008**, lancement du projet 1000 génomes pour établir le catalogue des variations génétiques humaines le plus détaillé. La phase pilote du projet est terminée, le séquençage de plus de 2000 génomes est en cours.
- Entre **2009 et 2010**, les chercheurs utilisent massivement le séquençage de la partie exprimée (fonctionnelle) du génome. Ce séquençage d'exomes (exon *versus* intron (intragenic region)) permet la découverte rapide des variants alléliques et des mutations.
- En **2012**, une annotation fonctionnelle du génome humain est publiée.
- En **2016**, la séquence de 10 000 génomes humaines set publiée avec un niveau élevé de couverture (30X à 40X).

b. Banques de données bibliographiques (PubMed, Google Scholar, Research Gate)

Ces banques de données permettent de trouver des articles scientifiques rapidement et manière intelligente.
Pour PubMed par exemple, nous pouvons chercher des sujets d'articles, des auteurs d'articles et trier selon nos besoins. Il est possible d'avoir les informations des auteurs et le texte complet pour chaque article. On peut également avoir le texte en PDF disponible au téléchargement.

c. Banques de données biologiques et leur utilisation

On remarque que le développement de la banque de données GenBank, hébergée par le NCBI, est exponentiel. En effet, on passe en dix ans de presque rien à 10 milliards de données pour atteindre 6 ans plus tard 70 milliards. De 1982 jusqu'à aujourd'hui, le nombre de données dans GenBank a doublé tout les 18 mois. Aujourd'hui on atteint des chiffres que l'esprit humain ne peut concevoir et acquérir.

Dans le Genome Project, on peut trouver des séquences nucléotidiques pour plus de 300 000 organismes avec des annotations bibliographiques et biologiques.

Un exemple de **système généraliste** : les bases de données du NCBI. GenBank est une banque de données de séquences biologiques (génomes/gènes, nucléotides, protéines). Si je cherche une protéine, je trouverais toutes fiches avec l'apparition de cette protéine.

https://www.ncbi.nlm.nih.gov/

Un exemple de **système spécialisé** : ENSEMBL (génomes)

Il est connecté au NCBI et on peut trouver des génomes et des chromosomes et leur localisation chez les espèces.

http://www.ensembl.org/index.html

Un autre exemple de système spécialisé : les banques de données taxonomiques (nécessaires pour rattacher une séquence biologique à l'organisme vivant dont elle est issue. Depuis le site de la NCBI on peut avoir des fiches taxonomiques sur les êtres vivants. Le site http://tolweb.org/tree est un site qui permet de retrouver le placement d'une espèce dans un arbre phylogénique.

d. Rechercher dans une banque de données de séquences biologiques

Identifier les séquences qui vont servir à faire un alignement multiple pour éventuellement reconstruire une histoire évolutive. Exemple : Utilisation de l'outil BLAST.

Comment trouver une séquence homologue à ma séquence requête parmi toutes les séquences connues ? Sachant qu'il y a 82 853 685 séquences…

➔Première solution : aligner toutes les séquences avec l'algorithme de comparaison de Smith et Watermann ou Needleman & Wunsch. L'avantage est de trouver la séquence la plus similaire. L'inconvénient est que cela dure 132 jours. La programmation dynamique donne un résultat exact.

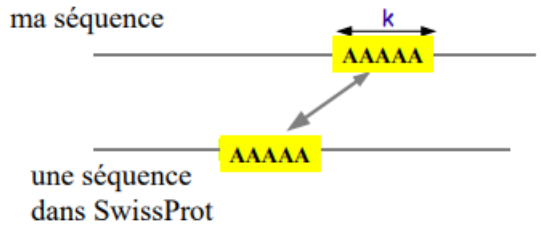
➔Deuxième solution : faire une pré-sélection sur les séquences et les aligner avec S&W. L'avantage est un résultat rapide et efficace. L'inconvénient est que l'on risque de louper la « perle rare ». C'est le processus qu'utilise BLAST, un processus **heuristique**.

L'**heuristique** (du grec *heuriskêin* « trouver ») est **l'utilisation de règles empiriques** pratiques, simples, rapides, facilitants la recherche des faits et l'analyse de situations soit dans un objectif de résolution de problèmes et de prise de décision soit dans un domaine particulier.

Les heuristiques trouvent leur place dans les algorithmes qui nécessitent l'exploration d'un grand nombre de cas car elles permettent de **réduire leur complexité moyenne en examinant d'abord les cas qui ont le plus de chances de donner la réponse.** Le choix d'une telle heuristique suppose de connaître déjà certaines propriétés statistiques sur l'ensemble d'instances du problème que l'on s'apprête à résoudre.

➔Alors comment pré-sélectionner les séquences ?

Supposons une séquence avec un **mot de longueur k** tel que :



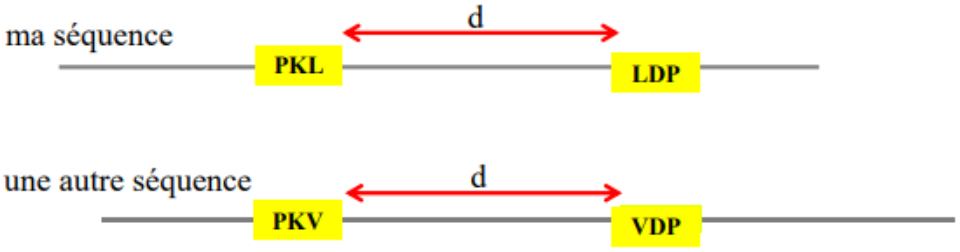
Première idée : ne retenir que les séquences partageant au moins un **mot de longueur k** avec ma séquence.

Cette présélection est rapide car toutes les séquences de SwissProt sont **indexées** c'est-à-dire qu'on détermine pour tous les mots de longueur k, la liste des séquences qui contiennent ce mot. SI l'on change la valeur de k, la présélection change.

Plus la taille k est grande, plus le critère est strict. Cependant, on risque de passer à côté de séquences homologues.
Si la taille k est trop petite, le critère n'est pas assez sélectif. On risque d'avoir trop de séquences candidates.

➔Comment faire ?

On applique un **double critère : les séquences doivent partager 2 petits mots au lieu d'un grand**. On peut faire trouver 2 paires de mots voisins à égale distance **d** dans les 2 séquences tel que :



Un premier score de similarité est établi.

Ensuite, on vérifie que l'on peut étendre ces amorces pour obtenir des alignements sans gaps de score de similarité.

Enfin on fait un alignement local Smith et Watermann.

La méthode S&W tel que 1 séquence contre tout UniProt dure 5 jours.
La méthode heuristique BLAST tel que 1 séquence contre tout UniProt dure quelques secondes.

Cours 4 : Le langage de l'ADN. Conservation, transmission et traduction de l'information génétique.
[RAPPEL COURS L1 S2 SCIENCES DE L'ENVIRONNEMENT]

Définition de l'ADN, réplication de l'ADN, reproduction de la cellule, mutations, expression du patrimoine génétique, transcription et traduction de l'information génétique et méthylation de l'ADN.

I. ADN, une molécule universelle et variable

Un chromosome comporte de nombreux gènes qui déterminent les caractères phénotypiques d'un individu. L'ADN ou acide désoxyribonucléique est le principal constituant des chromosomes et des gènes. La transgénèse est le fait de transférer un gène d'un organisme donneur à un organisme receveur. L'information portée par le gène est écrite dans un langage universel. L'universalité de l'ADN est un indice de la parenté de tous les êtres vivants.

Structure de l'ADN : deux chaînes ou brins de nucléotides enroulés en une double hélice.
Chaque nucléotide est constitué d'un groupement phosphate, d'un sucre (désoxyribose) et d'une base azotée

Adénine
Cytosine
Guanine
Thymine

Elles s'associent par des liaisons faibles par le principe de complémentarité des bases : A avec T et C avec G.Dans un gène c'est la succession ordonnée de nucléotides d'une chaîne d'ADN, ou séquence, qui constitue l'information génétique sous forme codée. Le squelette sucre-phosphate est constant.
Un même gène peut exister sous différentes versions : ce sont les allèles.
La biodiversité est composé de 3 niveaux : écosystème (milieu de vie (biotope), espèces (biocénose) et interactions), espèces, caractères (génétique et allélique)
La sélection naturelle est une variation non aléatoire de la fréquence des allèles.

II. Reproduction conforme de la cellule et réplication de l'ADN

- **L'interphase :** croissance cellulaire et réplication de l'ADN

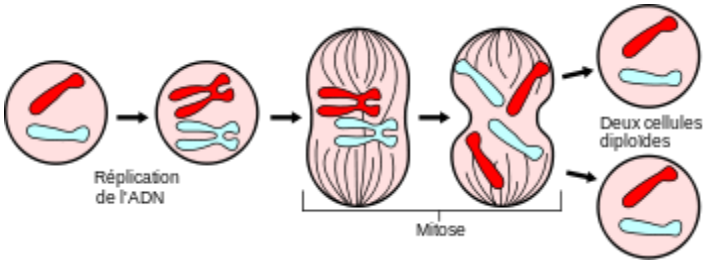
L'**interphase** est l'étape du cycle cellulaire précédant la mitose.
La phase **G1** est la phase de croissance cellulaire, les chromosomes à une chromatide sont invisibles.
La phase de synthèse **S** est le doublement de la quantité d'ADN
La phase **G2** est la préparation de la mitose, les chromosomes invisibles car non condensés, sont à deux chromatides
La phase **S** est la réplication semi-conservative de l'ADN d'une cellule préparant sa mitose.

1.Séparation des deux brins de la molécule d'ADN par une enzyme (**ADN polymérase**)
2.Copie de chacun des brins en ajoutant les nucléotides selon les règles de la complémentarité AT CG : la copie est donc conforme.
3.Chaque molécule d'ADN formée par réplication comprend un brin ancien et un brin nouveau qui a été polymérisé par l'ADN polymérase.
4.Les deux molécules d'ADN ainsi obtenues contiennent la même information génétique, aux erreurs de réplication près.
5.Ces deux molécules forment les deux chromatides identiques d'un chromosome à 2chromatides.

La réplication est l'étape qui permet de conserver la séquence nucléotidique car chaque brin d'ADN sert de modèle à la synthèse d'un nouveau brin.

- **La mitose** : conservation du caryotype

La mitose est une des étapes du cyle cellulaire qui se déroule en 4 phases :



Prophase : condensation de la chromatine du noyau, les chromosomes à deux chromatides deviennent visibles au microscope optique, l'enveloppe nucléaire disparait

Métaphase : Fin de la condensation des chromosomes à deux chromatides qui s'alignent, leurs centromères sont situés au niveau de l'équateur de la cellule, l'ensemble des chromosomes alignés forme la plaque équatoriale

Anaphase : Clivage des centromères de chaque chromosome et séparation des deux chromatides sœurs de chaque chromosome, chaque chromatide sœur migre vers un pôle de la cellule.
Télophase : Décondensation des chromosomes à une chromatide, l'enveloppe nucléaire se reforme dans chaque nouvelle cellule fille et le cytoplasme se sépare ; la cytodiérèse sépare les deux nouvelles cellules et marque la fin de la mitose.

La **mitose** est la division d'une cellule mère en deux cellules filles. La répartition égale des chromatides sœurs dans les 2 cellules filles : conservation des caractéristiques du caryotype= même nombre et même morphologie des chromosomes. Chaque chromatide sœur étant constituée d'une molécule d'ADN quasiment identique, les 2 cellules filles héritent de la même information génétique issue de la cellule mère. Pendant une mitose, la quantité d'ADN d'une cellule est divisée par 2, ce qui correspond à la séparation des chromatides de chaque chromosome.

III. Variabilité génétique et mutation de l'ADN

- Les mutations sont des modifications aléatoires de l'ADN

Les mutations sont des modifications spontanées et au hasard de la séquence d'ADN, elles peuvent survenir tout au long du cycle cellulaire.
Il y a trois types de mutations :

Substitution= changement d'un nucléotide en un autre (AGCT devient ATCT)
Insertion= ajout d'un ou plusieurs nucléotides (AGCT devient AGTTTCT)
Délétion= perte d'un ou plusieurs nucléotides (AGCT devient AT)

Lorsqu'il y a mutation, **une réparation est possible** par des **enzymes** : ces enzymes repèrent puis corrigent les dommages (nucléotides altérés, mésappariements) et certaines modifications échappent au système de réparation.
La mutation **peut être transmise** par mitose lorsqu'elle n'est pas léthale pour la cellule mère.
Les mutations peuvent affecter l'ADN d'une cellule **somatique** (mutation non héréditaire, non transmise à la descendance et qui disparaît à la mort de l'individu) ou d'une cellule **germinale** (qui se différencie en gamètes, mutation héréditaire qui peut se transmettre à la descendance).

Les **mutagènes** : éléments physiques ou chimiques modifiants/altérant l'ADN. Ex : benzène, ultraviolet, radioactivité… Les agents mutagènes **augmentent la fréquence d'apparition des mutation**s et augmente le risque de développer des cancers. Les mutations dans un gène entraînent l'apparition de nouveaux allèles. Ils sont sources de biodiversité génétique et sont un des moteurs à l'évolution.

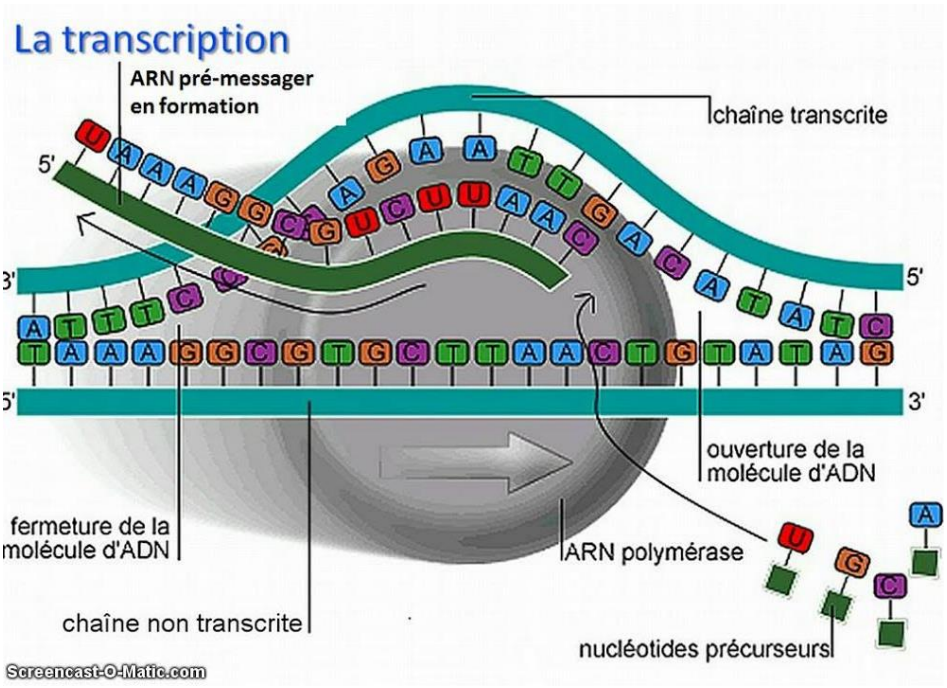
IV. L'expression du patrimoine génétique

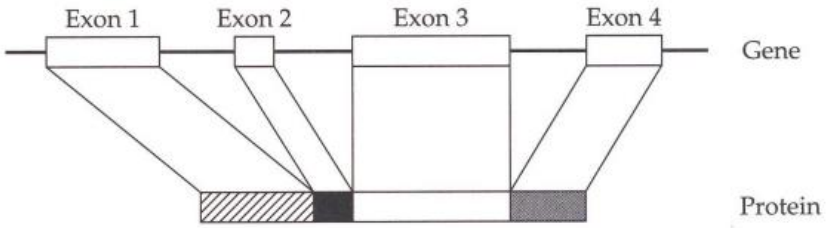
- La transcription, étape précoce de l'expression d'un gène

La **transcription** est la première étape de la synthèse des protéines. Elle se déroule dans le noyau. L'ARNpm (ARN pré-messager) est la séquence provenant d'un gène. Contrairement à l'ADN, c'est une molécule simple brin.

4 étapes :
Ouverture de la molécule d'ADN en séparant ses 2 brins au niveau des liaisons hydrogènes
Formation de la séquence d'ARN= ajout de ribonucléotides
Si G ADN alors C ARN et inversement
Si A ADN alors **U**racile dans l'ARN et si T ADN alors A ARN
La polymérisation des ribonucléotides entre eux permet de former la molécule d'ARNpm et de compléter la fermeture de la molécule d'ADN

La **maturation** de l'ARNpm est l'élimination de certains éléments (introns) et la conservation d'autres (exons). L'épissage se déroule dans le noyau et mène à la formation d'un ARNm. A partir d'un même ARNpm et selon l'épissage, **différents ARN messages** sont possibles.





<= Maturation

- La **traduction, seconde étape de l'expression d'un gène**

La traduction est la production d'une protéine à partir d'un ARNm, elle a lieu dans le cytoplasme et assure la correspondance entre l'ARNm, la séquence de nucléotides, et la protéine (séquence d'acides aminés). Le processus est réalisé par le ribosome selon le code génétique. **Un ribosome se place le long de l'ARN messager au niveau des codons (triplet de nucléotides) et glisse le long de l'ARN messager et synthétique la protéine, c'est l'élongation.**

Le code génétique associe **un codon de trois nucléotides avec un unique acide aminé**. Le code génétique est universel, redondant (plusieurs codons peuvent correspondre au même acide aminé) et univoque (un codon correspond à un seul acide aminé)
Deux types de codons non-associés à un acide aminé : **un codon initiateur** (codon AUG par lequel la traduction commence toujours) et **un codon stop** (met fin à l'élongation, donc à la traduction).
La polymérisation est la formation des liaisons peptidiques entre 2 acides aminés par le ribosome. Le premier acide aminé est la méthionine correspondant au codon initiateur AUG. Les acides aminés s'assemblent selon la séquence en acides nucléiques de l'ARN messager et la séquence nucléodique (ARN, donc ADN et gène) détermine la séquence peptidique de la protéine. La polymérisation s'arrête au codon stop, la protéine est libérée.

Le code génétique

Premier nucléotide	Deuxième nucléotide U		Deuxième nucléotide C		Deuxième nucléotide A		Deuxième nucléotide G		Troisième nucléotide
U	UUU	phényl-alanine	UCU	sérine	UAU	tyrosine	UGU	cystéine	U
	UUC		UCC		UAC		UGC		C
	UUA	leucine	UCA		UAA	STOP	UGA	STOP	A
	UUG		UCG		UAG		UGG	tryptophane	G
C	CUU	leucine	CCU	proline	CAU	histidine	CGU	arginine	U
	CUC		CCC		CAC		CGC		C
	CUA		CCA		CAA	glutamine	CGA		A
	CUG		CCG		CAG		CGG		G
A	AUU	isoleucine	ACU	thréonine	AAU	asparagine	AGU	sérine	U
	AUC		ACC		AAC		AGC		C
	AUA		ACA		AAA	lysine	AGA	arginine	A
	AUG	méthionine	ACG		AAG		AGG		G
G	GUU	valine	GCU	alanine	GAU	acide aspartique	GGU	glycine	U
	GUC		GCC		GAC		GGC		C
	GUA		GCA		GAA	acide glutamique	GGA		A
	GUG		GCG		GAG		GGG		G

SYNTHESE : Le passage du gène à la protéine se fait en deux étapes.
D'abord le segment de la molécule d'ADN correspondant au gène est copié sur **un brin d'ARN**, que l'on appelle ARNm, c'est la **transcription.**
Puis ce brin d'ARNm est à son tour recopié, mais dans un langage différent, **celui des acides aminés**, pour donner la séquence correspondant à la protéine synthétisée, c'est la **traduction**.
La séquence polypeptidique ainsi constituée prend alors une forme spatiale qui lui est spécifique pour devenir à proprement parler la protéine.
L'information génétique définit les fonctions biologiques d'un organisme (reproduction, nutriton, excrétion, action sur l'environnement, communication)

Un gène est une séquence d'acide désoxyribonucléique qui spécifie la synthèse d'une chaîne de polypeptides ou d'un acide ribonucléique fonctionnel. On peut également définir un gène comme une unité d'information génétique.

V. Epigénétique et méthylation

L'**épigénétique** est la discipline de la biologie qui étudie les mécanismes moléculaires qui modulent l'expression du patrimoine génétique en fonction du contexte.

La **méthylation** est l'ajout ou la substitution d'un groument méthyle (CH_3—. (radical alkyle hydrophobe).Un radical est une espèce chimique possédant un ou plusieurs électrons sur sa couche externe non appareillés. Un alkyle dérive un alcane par la perte d'un atome d'hydrogène. Hydrophobe signifie qui repousse l'eau. Le groupe méthyle est dérive du méthane CH_4.) sur un substrat. On trouve ce groupe méthyle dans des acides aminés codés par le code génétique lors de la traduction tel que l'alanine, la valine et la leucine.

La méthylation de l'ADN est un processus épigénétique dans lequel certaines bases nucléotidiques peuvent être modifiées par l'addition d'un groupement méthyle. Cette modification de l'Adn est effectuée par des enzymes particulières appelées DNMTs pour (DNA methyl-transferase).

La méthylation de l'ADN conditionne l'expression des gènes dans chaque cellule.

Chez les vertébrés, plusieurs types de nucléotides méthylés peuvent être rencontrés, principalement les cytosines et les adénines.

Prenons l'exemple de la méthylation de la cytosine chez les abeilles.

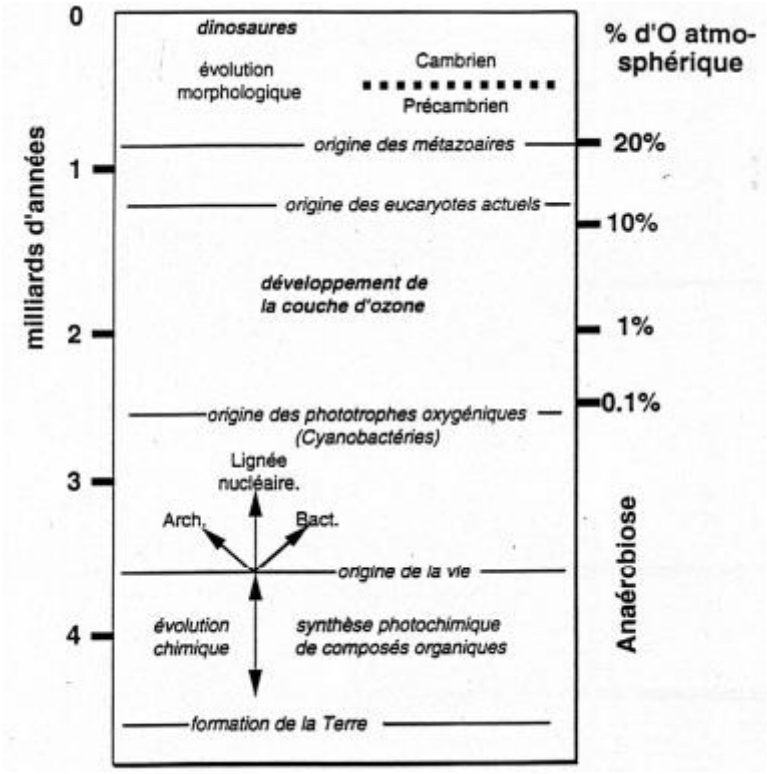
Une larve d'abeille nourrie 21 jours avec de la gelée royale devient une reine féconde vivant 3 à 5 ans. Une larve nourrie 3 jours avec de la gelée royale et 18 jours par du miel et du pollen devient un ouvrière stérile vivant quelques semaines. La gelée royale inhibe l'enzyme DNMT3 qui méthyle la cytosine de l'ADN.
Si il y a délétion du gène DNMT3, la cytosine de l'ADN est méthylée et on obtient que des reines.

Cours 5 : La réinterprétation des donnés et le changement de cadre conceptuel : Les bases de l'unité et de la diversité du vivant, concept d'espèce et arbres phylogénétiques.

Sommaire :

1. Bref aperçu de l'histoire des concepts en évolution. La nomenclature du vivant. Théories pour l'origine et l'évolution des espèces.
2. Notions de phylogénie et arbres phylogénétiques du vivant
    a. Utilisation de critères morpho-anatomiques
    b. Utilisation de critères de développement
    c. Utilisation de critères moléculaires
3. Que nous indique un arbre phylogénétique ?

I. Bref aperçu de l'histoire des concepts en évolution. La nomenclature du vivant. Théories pour l'origine et l'évolution des espèces.

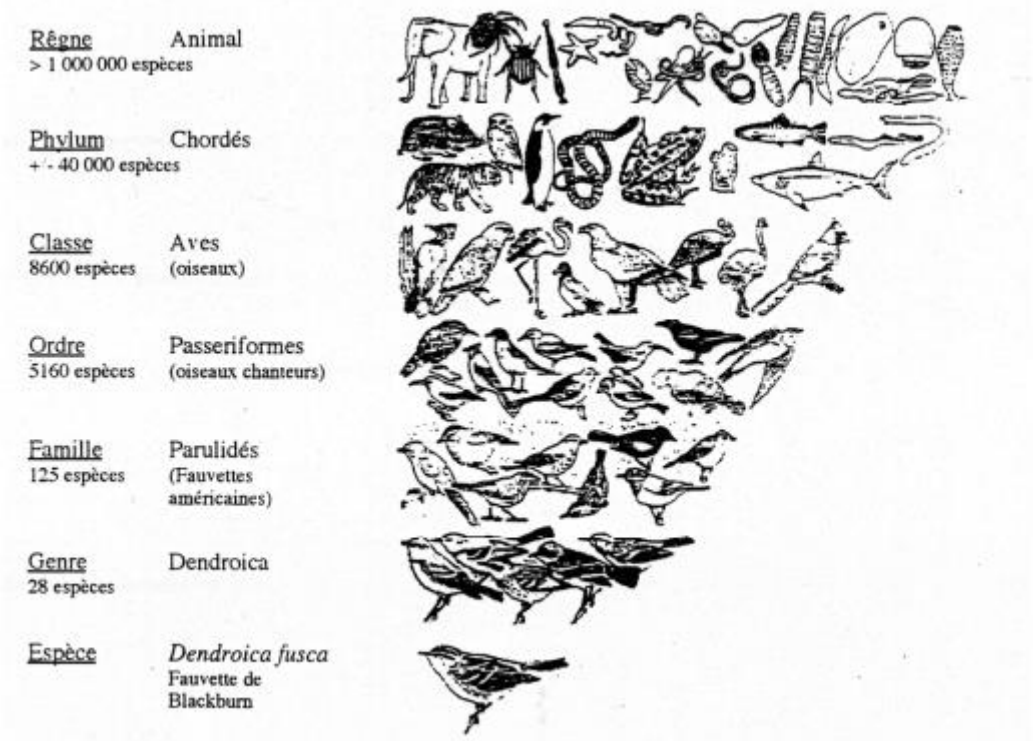


Il y a 4M d'années, suite à la formation de la Terre, une évolution chimique se forme et une synthèse photochimique de composés organique s'organise au fur et à mesure que l'oxygène augmente. Vers 3,7M d'années, on peut dater l'origine de la vie avec les bactéries, des lignées nucléaires (proto-eucaryotes), archées. Les cyanobactéries sont des phototrophes oxygéniques et apparaissent il y a 2,6M d'années (transformation du CO_2 en O_2). Entre 2,6 et 1,2M d'années, la couche d'ozone se développe et on passe de 0,1% à 10% d'oxygène atmosphérique. Les eucaryotes actuels apparaissent il y a 1,2M d'années. Puis viennent ensuite les métazoaires (ensemble des animaux, êtres eucaryotes, pluricellulaires, hétérotrophes et généralement mobiles) qui subissent des évolutions morphologiques. Il y'a 500 millions d'années, les eucaryotes multicellulaires se diversifient en plantes, champignons, animaux…etc.

- Concepts avant le 19ème siècle pour l'origine des espèces

L'origine et l'histoire des êtres vivants sur Terre fait l'objet de nombreuses superstitions, spéculations philosophiques et visions religieuses (« théologie naturelle » ou « créationnisme » soit le courant théologique qui consiste à « connaître Dieu à partir de l'expérience du monde » et Dieu comme origine de tout.

- Dans son livre *Histoire des animaux*, Aristote (384-322 av. J.C.) fait une classification en 3 catégories : *Enaima* (avec sang) *Anaima* (sans sang (insectes, mollusques)), *Entoma* (corps divisé en segments (généralement 3 parties, abdomen thorax et tête) par incisions (*entomos*). *Entoma* fut traduit par *insecta* en latin. Il établit une *scala naturae* (échelle naturelle) avec l'être humain tout en haut, la matière animée et la matière inanimée en bas (métaux, sels et roches).
- Pline l'Ancien (23-79) sort les 37 tomes de son *Histoire naturelle* qui contient de la biologie, zoologie, géologie.
- Albert le Grand (1193-1280) et son *De animalibus*, une étude zoologique en 26 volumes qui ajoute une description sur le comportement des animaux, leur mode de reproduction et leur anatomie. C'est un commentaire de l'ouvrage d'Aristote.
- Au XVIème siècle, Conrad Gesner (1516-1565) fait une classification illustrée des animaux dans *Historia animalium*.
- Pour l'Archevêque James Ussher (1664) pense que la Terre a été créée par la Trinité le 26 octobre 4004 av. J.C. (calcul effectué par l'Ancien Testament).
- **Carl Von Linné** (1707-1778) : Il publie le *Systema naturae* (système morpho anatomique). Selon lui, il n'y a pas d'explication aux ressemblances entre espèces, c'est l'œuvre de Dieu. Les espèces sont des formes similaires mais disjointes les unes des autres. En 1735, il publie le premier essai de classification systématique des trois règnes minéral, végétal et animal. Il divise les animaux en six groupes (quadrupèdes, oiseaux, amphibiens, poissons, insectes et vers) en fonctions d'organes spécifiques (dents, bec, nageoires ou ailes). Lors de la dixième édition, en 1758 il généralise le système de nomenclature binomial. Il veut imposer un principe descriptif rationnel et universel. ⬇Nomenclature de Linné⬇



- **Georges-Louis Leclerc de Buffon** (1707-1788) est un naturaliste, mathématicien, philosophe, écrivain, biologiste et cosmologiste français. Dans son *Histoire naturelle* il établit un lien entre le singe et l'homme, le singe étant un « homme dégénéré », il fait une classification transformiste. Il fait aussi des travaux dans la datation de la Terre, l'homme et les animaux : il approche la notion d'évolution par celle de « dégénérescence ». Les espèces peuvent

changer selon leur environnement. Il contribue aussi à des connaissances en géologie en introduisant la notion de bio géographie et de chronologie de la Terre. Il fait des travaux sur la datation de la Terre.

➔Comment définir une espèce ?

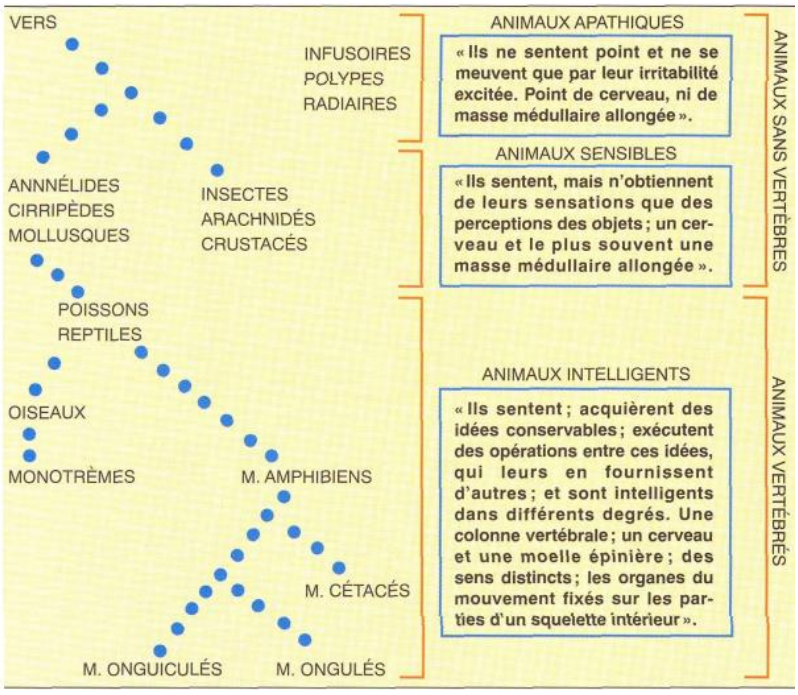
C'est une communauté d'êtres vivants **interféconds** pouvant échanger du matériel génétique et produisant des descendants eux-mêmes **féconds** (en effet, certains individus du même genre mais appartenant à des espèces différentes peuvent se croiser pour donner un individu hybride mais celui-ci est le plus souvent stérile)
Dans le cas de la multiplication strictement asexuée, on parle d'espèces à la place de lignée, le groupement étant alors purement phylogénétique.

➔Combien d'espèces ?

Le règne animal possèderait 7 770 000 espèces. Sur Terre, on a 1,244,360 espèces cataloguées et on prédit 8,750 000. Dans les océans, on aurait 194,409 espèces cataloguées pour une prédiction de 2,210 000 espèces.

- Les précurseurs des théories évolutionnistes du XIXème siècle
    - Lamarck et l'hérédité des caractères acquis

Jean Baptiste de Lamarck (1744-1829) est l'inventeur des termes « biologie, invertébrés, vertébrés, insectes ». C'est un théoricien de l'évolution. Il est partisan de la théorie du **transformisme**.
L'environnement agit directement sur les organismes vivants et provoquent leur transformation.
L'usage (développement) ou non (atrophie) des organes est le facteur clé expliquant leur modification dans le temps. Il théorise **l'hérédité des caractère acquis** : Les traits acquis sont transmissibles à la génération suivante. De nouvelles espèces naissent à partir d'une certaine accumulation de modifications. Par exemple la girafe a eu besoin d'avoir un cou plus long pour manger les feuilles des arbres et a acquis ce caractère par la suite. La taupe a perdu l'usage des yeux par défaut constant d'exercice, puisque la taupe vit sous terre.



Dans *Philosophie Zoologiques* (1809), Lamarck établit un tableau sous forme d'arbre qui schématise l'origine des différents animaux. Il y aurait génération spontanée au niveau des vers, qui ensuite se complexifient continuellement. La matière inanimée se complexifie jusqu'à devenir des êtres complexes.

Un des problèmes de la théorie de Lamarck est celui de la non prise en compte de la génétique et du facteur hasard que la génétique induit.

- Cuvier et le fixisme

Georges Cuvier (1769-1832) est un anatomiste étudiant les fossiles. Il est opposé à l'évolutionnisme et à la théorie de Lamarck. Pour lui, les espèces vivantes restent immuables, c'est la position du **fixisme**. De nouvelles formes de vie peuvent apparaître et d'autres disparaître à l'occasion de périodes de destruction, c'est la position du **catastrophisme**.

- Darwin et la sélection naturelle

Charles Darwin (1809-1882) est un naturaliste paléontologue britannique du XIXème siècle. Il a révolutionné la biologie animale et végétale par ses nombreux écrits. Il a contribué à de nombreuses avancés dans tous les domaines des sciences de la vie. Il se lance dans l'étude de la géologie et persuade son père pour le laisser partir en tant que volontaire naturaliste sur le Beagle pour aider le Capitane Fitzroy dans des travaux de géologie. Il fait un compte rendu de ce voyage dans un journal de bord, *Voyage du Beagle* où il fait des observations naturalistes en Afrique, en Australie et en Amérique du Sud (notamment dans les îles Galapagos). Il ajoute que les espèces existantes ne sont pas fixes et qu'elles se modifient. De retour en Angleterre, il classe, documente et publie ses découvertes et entretient des correspondances avec des savants du monde entier pour plus de 20 ans de réflexions. Il reçoit un manuscrit d'Alfred Russel Wallace qui lui donne l'idée de transformation des espèces par le jeu de la sélections naturelle. Son ouvrage *L'origine des* espèces ou de son titre original *On the origin of species by means of natural selection or the preservation of favoured races in the struggle for life* sort en Novembre 1859 et c'est un grand succès. Il est très critique sur son ouvrage qui pour lui manque de clarté et pense que ses capacités sont moyennes. Il s'intéresse aussi dans ce livre aux plantes et aux émotions chez les hommes et animaux. Il est très critiqué et est sujet de beaucoup de caricatures.

« Toutes les parties de l'organisation et tous les instincts offrent au moins des différences individuelles ; la lutte constante pour l'existence détermine la conservation des déviations de structure ou d'instinct qui peuvent être avantageuses ; et enfin, des gradations dans l'état de perfection de chaque organise, toutes bonnes en elles-mêmes, peuvent avoir existé. Je ne crois pas que l'on puisse contester la vérité de ces propositions. » « La nature ne fait pas de sauts »
➔Dans ce paragraphe Darwin pose le principe **que la sélection naturelle favorise dans la reproduction,** les individus qui possèdent des caractéristiques avantageuses dans leur environnement. **Une espèce évolue au cours du temps et avance par strates.** Darwin transfère ici à la biologie la théorie de l'économiste Malthus avec le « combat pour l'existence »

« Nous savons aussi qu'un croisement entre des individus distincts de la même variété, et entre des individus appartenant à des variétés différentes augmente le nombre des descendants, et augmente certainement leur taille ainsi que leur force. Cela résulte principalement du fait que les formes que l'on croise ont été exposées à des conditions d'existence quelque peu différentes. »
➔ Darwin prolonge l'argument de la sélection naturelle par le biais de la reproduction : **les individus les plus avantagés sont plus propices à survivre et à être sujet à reproduction que les autres**, la descendance héritera de ces caractéristiques avantageuses. Darwin fait évoluer le transformisme de Lamarck par le principe de **macroévolution** (variations entre êtres vivants à l'origine des espèces). La sélection naturelle favorise les formes les mieux adaptées.

« La théorie de la sélection naturelle impliquant l'existence antérieure d'une foule innombrable de formes intermédiaires reliant, les unes aux autres, par des nuances aussi délicates que le sont nos variétés actuelles, toutes les espèces de chaque groupe, on peut se demander pourquoi nous ne votons pas autour de nous toutes ces formes intermédiaires et pourquoi tous les êtres organisés ne sont pas confondus en un inextricable chaos »

➔Dans ce paragraphe, Darwin établit **que l'ensemble des êtres vivants ne forme pas un chaos mais des chaînes d'évolutions par strates intermédiaires que l'on classer et penser**. Il pense une classification ramifiée. Toutes les espèces ont un ancêtre commun, et au fil du temps apparaissent des variations entre êtres vivants à l'origine des espèces. C'est la fin du créationnisme.

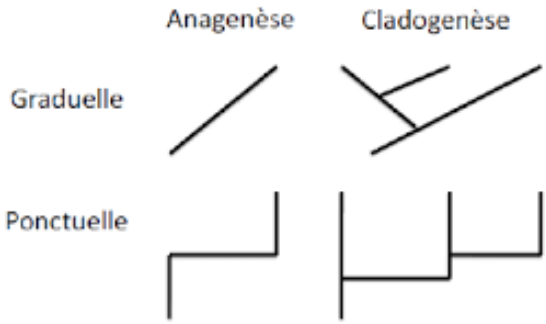
« Ce sont les espèces dominantes et les plus répandues qui varient le plus et le plus souvent, et les variétés sont souvent locales ; or, ce sont là deux circonstances qui rendent fort peu probable la découverte de chaînons intermédiaires dans une forme quelconque. Les variétés locales ne se disséminent guère dans d'autres régions éloignées avant de s'être considérablement modifiées et perfectionnées ; quand elles ont émigré et qu'on les trouve dans une formation géologique, elles paraissent y avoir été subitement créées, et on les considère simplement comme des espèces nouvelles. »
➔Darwin intègre **la mobilité des êtres vivants** dans le critère d'évolution : une fois que les individus se sont perfectionnés au maximum ils peuvent **migrer** ensuite.

Si l'on fait la synthèse des théories de l'évolution, on peut représenter **l'évolutionnisme** sous forme d'arbre (les espèces ont une origine commune et se modifient au cours du temps), **le transformisme** comme des diagonales (les espèces se transforment à partir d'origine distinctes) et **le fixisme** comme des droites discontinues.

Evolutionnisme **Transformisme** **Fixisme**

- Les principes **d'anagenèse et de cladogenèse**

Bernhard Rensch (1900-1990) est un biologiste allemand ayant fait partie d'un groupe scientifique qui a élaboré la théorie **synthétique** de l'évolution au cours des années 1930/40. C'est une théorie darwinienne de l'évolution basée sur la sélection naturelle et les variations du génome basées sur le hasard. C'est une synthèse des diverses théories biologiques du XIXème siècle et du XXème siècle.



Rensch popularise le terme d'**anagenèse** (aussi appellée évolution phylétique).
C'est « l'évolution temporelle et graduelle d'une lignée unique qui se transforme progressivement pour donner naissance à une nouvelle espèce, sans ramification. Une espèce descendante remplace une espèce ancestrale dans la continuité formant une lignée évolutive »

La **cladogenèse** (terme aussi popularisé par Rensch) est « le processus évolutif aboutissant à l'apparition de nouvelles espèces par la scission de nouvelles espèces par la scission d'une lignée ancestrale en au moins deux lignées descendante. Chacun des deux rameaux évolue alors de manière isolée et peut donner naissance à une nouvelle

espèce. La **spéciation** par cladogenèse est la conséquence de phénomènes de mutations, sélections et ségrégations plus ou moins progressifs, sous l'effet d'une cause externe ou interne.

Deux types de cladogenèse :

Ponctué : une vaste population subit une séparation (Amérique du Sud et Afrique il y a 110 millions d'années par exemple) tel qu'une petite population marginale et le reste de la population vont se retrouver isolés géographiquement. Du fait du faible effectif, elle évolue rapidement par dérive génétique sur son territoire. Lorsque l'isolement géographie se termine, les deux populations sont suffisamment différentes pour ne plus se croiser. Il y a donc un isolement reproducteur qui est synonyme de la création d'une nouvelle espèce

Graduel et phylétique : une vaste population subit une séparation en deux populations importantes et un isolement géographique vont diverger graduellement car elles sont soumises à des forces sélectives différentes. Des caractères nouveaux apparaissent et envahissent chaque ensemble de population, tout ceci va produire alors une barrière reproductive pour produire deux nouvelles espèces.

- **L'hypothèse de la reine rouge**

Leigh Van Valen propose dans un article de 1973 que la complexification du vivant est une course entre les êtres vivants où chaque espèce reste en deçà de son optimum adaptatif en raison innovations continuelles de ses compétiteurs. Chaque espèce tends à évoluer par concurrence avec les autres espèces. Les espèces ne sont pas rendues « meilleures ».

Cette hypothèse est nommée Reine Rouge en référence au personnage de la Reine Rouge dans *De l'autre côté du miroir* (Lewis Caroll), personnage qui entraîne Alice dans un miroir dans une course immobile. Alice s'étonne de la course car elle pense que si l'on court on arrive forcément quelque part. La reine répond que le pays d'origine d'Alice est lent. Alice reprend : « Tandis qu'ici il faut courir de toute la vitesse de ses jambes pour simplement rester là où l'on est. Si l'on veut aller quelque part, ailleurs [évoluer donc], il faut courir au moins deux fois plus vite que ça ! ».

II. Notions de phylogénie et arbres phylogénétiques du vivant

a. Utilisation de critères morpho-anatomiques

La classification linnéenne, inspirée de l'inventaire du vivant proposé par Carl von Linné (1707-1778) et élaborée dans le cadre du fixisme, procède par coupure, par séparation, et non par regroupement.
Dans un traité morpho-anatomique, on classe les espèces à partir d'un critère morpho-anatomique (par exemple le carène pour les oiseaux) et on fait des classes et des regroupements.

C'est une clé simplifiée de systématique. On part de critères généraux (polarité, yeux, bouches, corps…) pour aller à du plus spécifique. Selon la présence ou non de ces caractères on range les espèces.

On utilise des critères morphologiques (voire géographiques et écologiques) pour identifier un animal. Les critères anatomiques, physiologiques, embryologiques, biochimiques et génétiques ne sont pas considérés. Lorsqu'un caractère morphologique est pris en compte, soit le spécimen à identifier présente ce caractère morphologique, soit le spécimen à identifier ne présente pas ce caractère morphologique.

On retrouve alors un système de répartition normalement **dichotomique.**

Les critères morphologiques utilisés au départ doivent permettre de distinguer « facilement » les **principaux taxons.** Ces critères sont ensuite de plus en plus détaillés pour différencier les subdivisions jusqu'à correspondre à **une et seule espèce**.

Cette clé simplifiée ne dépasse pas les taxons inférieurs sauf pour les cordés.

Le système n'est pas strictement dichotomique. Par exemple, le **niveau 1** (polarité) propose 4 options (pas de polarité visible, un seul axe, deux axes et trois axes).

Ce système aboutit à une **arborisation** dont l'agencement des embranchements ne correspondra pas à celui présenté lors des premières séances de TD. Les caractéristiques des taxons restent cependant identiques. Toute la subtilité réside dans l'appréciation du vocabulaire utilisé et la connaissance des structures morphologiques évoquées.

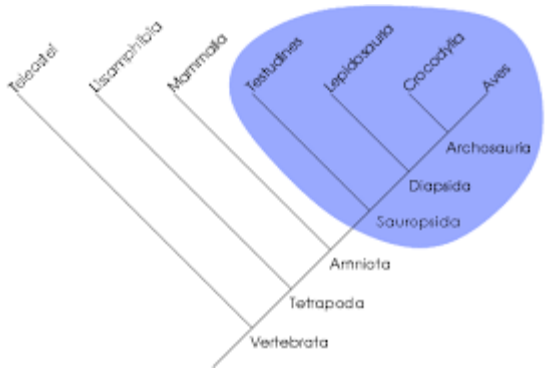
Au contraire, la systématique phylogénétique, aussi nommée **cladistique**, a rompu avec les procédures de séparation, pour se concentrer sur des regroupements justifiés par ce que les membres d'un ensemble partagent vraiment. **Conséquences** : abandon des invertébrés, prosimiens, reptiles, poissons et nombre des groupes linnéens, pour d'autres dont l'agencement dans l'arbres respecte l'apparentement.

Darwin a fait un arbre théorique où il classe les espèces en vertu de souches communes. Hennig expose lui une méthode empirique qui permet de faire le classement de Darwin. Tandis que Darwin expose une généalogie théorique, Hennig expose la phylogénie. Dans le cadre empirique, l'arbre classe selon « qui est apparenté à qui ».
Dans un système cladistique, on établit des correspondances selon les formes d'organismes, les squelettes, les plans d'organisations des êtres vivants.

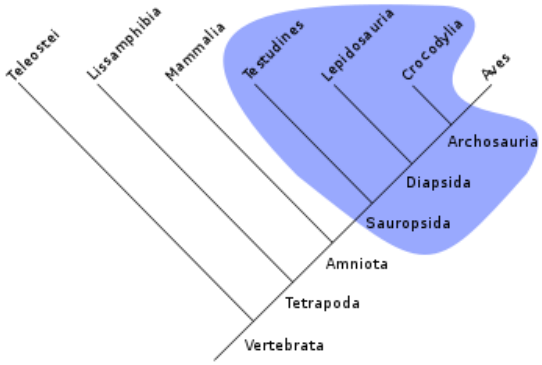
La **parcimonie** est le principe de simplicité ou principe d'économie ou le principe selon lequel « les hypothèses suffisantes les plus simples sont les plus vraisemblables ». En science et en philosophie, la **parcimonie** est un principe consistant à n'utiliser que **le minimum de causes élémentaires pour expliquer un phénomène.** Par exemple en **génétique**, ce principe privilégie **la séquence avec le minimum de changements évolutifs pour expliquer des relations phylogénétiques.**

Ainsi, lors de la classification évolutive de plusieurs taxons selon plusieurs caractères (par exemple avec un arbre phylogénétique), on applique **le principe de parcimonie en cas de difficulté à déterminer quelles espèces sont les plus proches** : l'hypothèse de phylogénie retenue est celle qui **maximise le nombre d'hypothèses d'homologie** (similitude due à une ascendance commune) tout en **minimisant le nombre d'hypothèses d'homoplasie** (similitude due à une adaptation convergente).

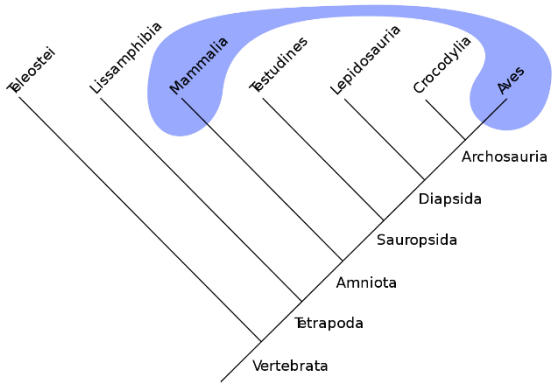
La **monophylie** est la caractéristique d'un groupe qui contient l'espèce souche dont descendent tous ses membres. ⬇



La **paraphylie** est le fait de mettre dans un même groupe taxinomique une espèce ancestrale et une partie seulement de ses descendants.⬇



La **polyphylie** est le fait de mettre dans un même groupe des taxons en excluant leur ancêtre commun.⬇



b. Utilisation de critères de développement

On peut utiliser la notion d'homologie entre les êtres vivants pour voir comment certains critères se sont développés avec le temps.

L'**embryologie comparative** permet de comparer les embryons de plusieurs espèces. Elle donne en faveur de l'évolution car il y a des similitudes développementales entre différentes espèces.

L'embryon entre l'homme le poulet, le serpent, la salamandre, le poisson est à peu près le même avec la présence d'une queue. **Deux structures sont homologues si elles sont une même origine embryonnaire.**

c. Utilisation de critères moléculaires

Evolution des molécules, évolution des gènes, évolution des génomes➔ évolution des espèces. On peut établir des parentés grâces aux acides aminés et aux codons pour voir les ressemblances

L'ARN ribosomique est une bonne molécule pour faire l'arbre phylogénétique de l'ensemble des êtres vivants car elle est aussi présente chez les bactéries.

III. Que nous indique un arbre phylogénétique ?

Un **arbre phylogénétique** est littéralement l'arbre qui raconte la genèse des lignées. « Phylogénie » vient du grec *phulon* (lignée) et *genesis* (genèse). Historiquement ce terme n'a pas été défini par rapport aux gènes.

La **phylogénie** (phylogenèse) est une reconstruction de l'histoire évolutive des êtres vivants. Le terme phylogenèse a été introduit par Haeckel en 1866. Les fondements de la systématique phylogénétique ou cladistique ont été formulés par Willi Hennig en 1950.

L'analyse phylogénétique a pour objectif de reconstruire des liens de parenté entre les organismes et d'estimer leurs temps de divergence.

Un **arbre phylogénétique** est une représentation graphique de la phylogenèse d'un groupe de taxons. Les sommets représentent les taxon ou les unités évolutives (OUTs-operational taxonomix units). Les nœuds internes représentent des ancêtres hypothétiques. Les branches définissent les relations entre les taxons en terme de descendance.

Il existe plusieurs types d'arbres (dendrogrammes) selon les méthodes avec lesquelles ils ont été construits :

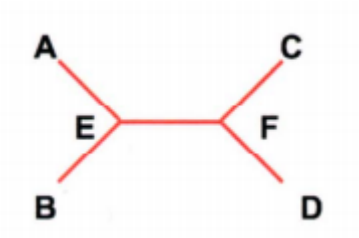
- **Phénogramme** : un dendrogramme obtenu par méthodes de distance où les relations entre taxa expriment les degrés de similitude globale.
- **Cladogramme** : un dendrogramme exprimant les relations phylogénétiques entre taxa et construit à partir de l'analyse cladistiques.
- **Phylogramme** : un cladogramme dont la longueur des branches est proportionnelle au nombre de changements évolutifs.

La longueur d'une branche est proportionnelle à la distance évolutive entre les séquences et leurs ancêtres (unité= substitution/site). Une branche est un trait et une feuille est le nom des espèces.

La plupart des méthodes de phylogénie produisent des arbres sans racine (unrooted tree *versus* rooted tree). Ceci est du au fait que ces méthodes détectent les différences entre les séquences, mais n'ont pas les moyens d'orienter ces changements par rapport au temps.

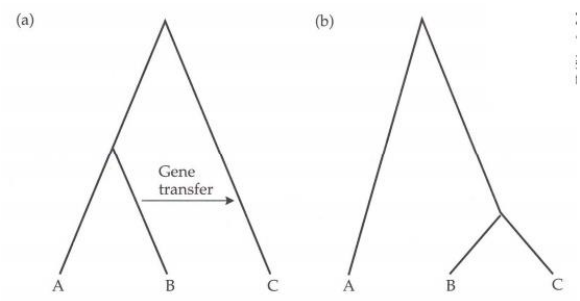
Il existe deux moyens pour enraciner un arbre sans racine :

- La méthode du **groupe extérieur** : il faut inclure dans l'analyse une séquence ou un groupe de séquence dont on postule *a priori* qu'ils sont externes au groupe étudié ; la racine est par nécessité sur la branche qui rejoint le groupe extérieur avec les autres séquences.
- Faire l'hypothèse de **l'horloge moléculaire** : toutes les lignées sont supposées avec évoluées avec la même vitesse depuis leur divergence avec leur ancêtre commun. La racine est située au point équidistant de toutes les feuilles de l'arbre.



Un **arbre non enraciné** est une représentation intemporelle des relations phylogénétiques.

Un **arbre enraciné** spécifie où se situe l'ancêtre commun de tous les taxons présents dans l'arbre, considéré comme la racine de l'arbre. L'arbre enraciné est orienté dans le temps.



L'histoire évolutive des gènes révèle celle des espèces qui portent ces gènes, à l'exception de la présence de **transfert horizontal** (transfert de gène entre espèce) et de **duplication génique** (orthologue/paralogue)

Remarques sur la représentation génétique :

Il n'y a pas de hiérarchie objective dans les taxons existant actuellement. Toutes les espèces sont évoluées, et présentent chacune une mosaïque d'états dérivés et d'états ancestraux de caractères, mosaïque variable d'un taxon à l'autre.

Cours 6 : Diversité génétique et code barre moléculaire.

I. Diversité génétique, sélection et reproduction

a. Cycle chromosomique, cycle de vie, reproduction sexuée et méiose

La **reproduction sexuée** implique un échange de patrimoine génétique entre les individus d'une **même espèce.**

➔ Qu'est-ce que le sexe ?

Ce sont les **échanges (ou recombinaison génétiques)**, la production d'une nouvelle information composite à partir **de deux ou plusieurs autres.**

C'est la notion de **parasexualité,** les échanges génétiques **sans reproduction d'individus.**

La reproduction sexuée constitue **les échanges génétiques** associés **à la reproduction des individus.**

La **parthénogenèse** est la division à partir d'un gamète femelle non fécondé. C'est un mode de reproduction monoparental comme l'auto-fécondation qui nécessite quant à elle l'intervention des deux gamètes, mâles et femelles, apportés par le même individu hermaphrodite. On parle de parthénogenèse **thélytoque** lorsque les mâles sont absents, la descendance est uniquement composée de femelles. On parle de parthénogenèse **arrhénotoque** quand les femelles sont absentes, la descendance est uniquement composée de femelles. On parle de parthénogenèse **deutérotoque** quand il y a parfois apparition de mâles de manière cyclique à une période de l'année. Il y a production de mâles et de femelles.

La **méiose** est le processus de double division cellulaire permettant la formation de gamètes, ou cellules sexuelles chez les organismes eucaryotes. Elle se distingue de la **mitose**, division cellulaire **somatique** qui ne se déroule qu'en un temps. A partir d'une cellule mère **diploïde** on obtient quatre cellules filles **haploïdes** au matériel génétique différent. Il n'y a pas de duplication de l'ADN lors de la méiose.

Une première phase appelée « **interphase** » avant la méiose va permettre la duplication de l'ADN et la condensation en chromatides. Les chromosomes **homologues** vont s'enjamber en certains points et échanger du matériel génétique. La cellule est dite « **diploïde** », elle présente deux exemplaires de chaque chromosome. Les chromosomes homologues vont se retrouver séquestrés sur un plan équatorial puis séparés. Chacun va migrer vers un des deux pôles de la cellule. Il y a alors un premier **brassage génétique** (ensemble des recombinaisons génétiques dans une population)**, le brassage interchromosomique.** On va avoir alors deux premières cellules filles dites « **haploïdes** »(elles ne contiennent qu'un exemplaire de chaque chromosome) lors de la première méiose.

La deuxième méiose va donner à partir des deux premières cellules filles, quatre cellules filles. On va avoir une autre division et les chromatides vont être séparées pour se retrouver dans une nouvelle cellule appelée **gamète**. Ce processus génère de la **diversité génétique** car cela permet d'aboutir à la formation de cellules haploïdes conçues pour fusionner avec d'autres pour obtenir des cellules diploïdes après fécondation.

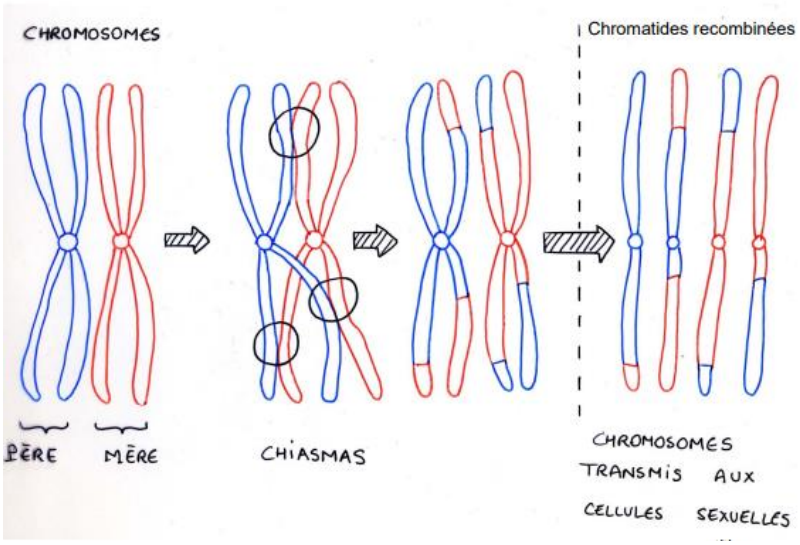
En effet, un **ovocyte haploïde** et un **spermatozoïde haploïde** va former après fécondation un **zygote diploïde.**

Il y a différents types de cycle chromosomiques chez les êtres vivants :

Les cycles **haplodiplophasiques** sont des cycles que l'on retrouve chez de nombreux végétaux avec une équilibre des deux phases haploïde et diploïde.
Les cycles **diplophasiques** avec prédominance de la phase diploïde que l'on retrouve chez de nombreux animaux et fucus.
Les cycles **haplophasiques** avec prédominance de la phase haploïde que l'on retrouve chez les spirogyre.

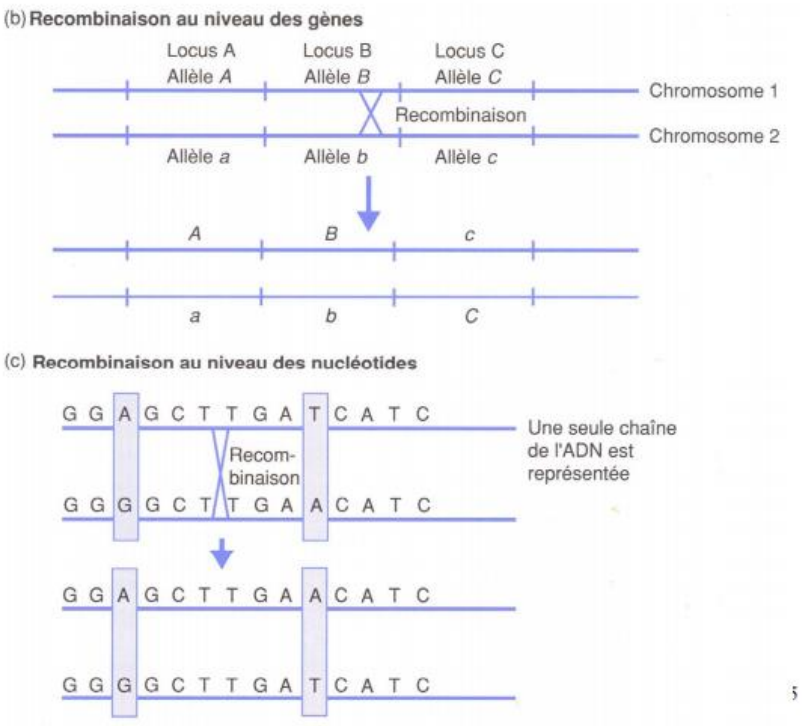
Lors des mitoses successives suite au zygote, le zygote va donner naissance à des cellules immortelles qui vont-elles mêmes donner des cellules immortelles dans le cadre d'une **lignée germinale** (héréditaire) et des cellules mortelles dans la cadre d'une **lignée somatique** (non héréditaire). La cellule va donc exprimer des fonctions conférant le caractère mortel de la cellule.
Les cellules germinales se trouvent dans les ovaires et testicules, les cellules somatiques constituent le reste des cellules du corps.

La méoise permet donc de former des gamètes haploïdes. La fécondation réunit deux gamètes haploïdes pour former une cellule-œuf diploïde. **L'alternance méoise/fécondation permet donc de maintenir la stabilité du caryotype dans l'espèce humaine**. Le principe est le même chez les autres espèces animales.



Lors de la méiose, le brassage génétique interchromosomique s'effectue par un procédé de **crossover ou de chiasme** tel que l'on va effectuer une recombinaison des chromatides.

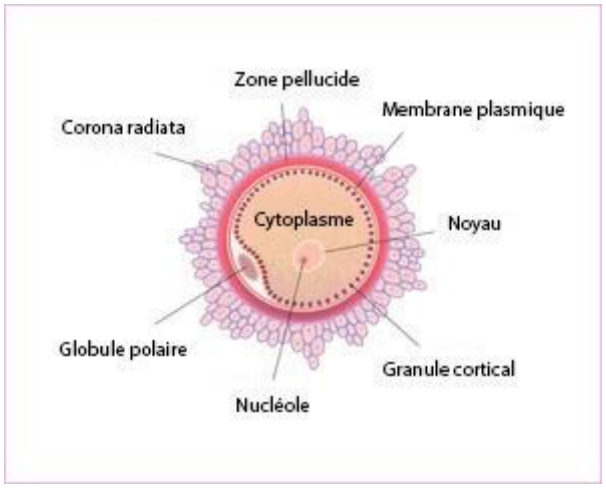
Cette recombinaison se produit avec des protéines d'assemblage et un système de synapse et se termine avec des protéines de désassemblage

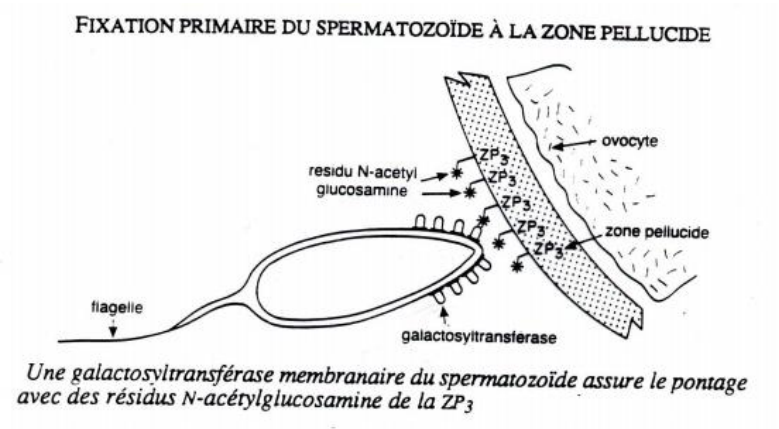


Les recombinaisons se font au niveau des gènes et des nucléotides tel que ←

Au niveau de la fécondation, l'ovule est constitué d'une zone pellucide, d'une membrane plasmique, de granules cortical, d'un noyau avec un nucléole à l'intérieur, d'un cytoplasme et d'un globule polaire.

Le spermatozoïde va utilise une protéine, le **galactosyltransférase membranaire** pour se fixer sur la zone pellucide de l'ovule qui contient des résidus N-**acétylglucosamine** de la macromolécule **ZP_3**.





La **diversité génétique** entre individus d'une même espèce se fait par des différences génétiques entre individus d'une même espèce qui apparaissent au fil du temps du fait des **mutations** (**modifications de séquences transmises aux descendants (microévolution)**. Le génome est à 99,9% identique entre les diverses populations du globe. On définit une **population** comme un groupe à base géographique, culturelle…au sein duquel se fait la reproduction (pool génétique). On parle **d'endogamie** quand les membres d'un groupe social choisissent leur conjoint à l'intérieur de ce groupe. On parle **d'exogamie** quand les membres d'un groupe social choisissent leur conjoint à l'extérieur de ce groupe.

Une **race** (définition du Larousse) est une population animale résultant, par sélection, de la subdivision d'une même espèce et possédant un certain nombre de caractères communs transmissibles d'une génération à la suivante.

C'est la catégorie de classement de l'espèce humaine selon des critères morphologiques ou culturels, sans aucune base scientifique et dont l'emploi est au fondement des divers racismes et de leur pratiques. (Face à la diversité humaine, une classification sur les critères les plus immédiatement apparents [couleur de la peau surtout] a été mise en place et a prévalu tout au long du XIXème siècle. Les progrès de la génétique conduisent aujourd'hui à rejeter toute tentative de classification raciale chez les êtres humains.

Littéraire : Lignée familiale considérée dans sa continuité ; ensemble des ascendants ou des descendants d'un personnage ou d'un groupe humain.
Ensemble de personnes présentant des caractères communs (profession, comportement, etc.), et que l'on réunit dans une même catégorie.

b. Sélection naturelle et reproduction sexuée

i. La sélection porte sur les gènes

La sélection naturelle porte sur l'information, les gènes. Les gènes que l'on observe aujourd'hui sont ceux **qui se sont le mieux répliqués lors des générations précédentes**. Ce qui est transmis d'une génération ce sont les informations codant le développement des organismes : la sélection opère un tri parmi ces informations, ne conservant que celles qui ont les meilleures propriétés de réplication. Les individus sont des entités transitoires, ce qui persiste de génération en génération, ce sont les gènes et non pas l'espèce.

« Les organismes sont des artifices inventés par les gènes pour les reproduire », Richard Dawkins, *Le gène égoïste*

ii. Sélection et compétition

La sélection naturelle ne favorise pas la « survie de l'individu » ou « la survie de l'espèce » : la sélection naturelle favorise les gènes qui permettent **une meilleure appropriation des ressources que les autres**.
La première idée intuitive est que la sélection naturelle favoriserait la survie de l'espèce. Tel que la sélection naturelle garde seulement les gènes qui maximisent la reproduction des individus (en négligeant tous les systèmes « parasites ». Donc grâce à la sélection naturelle, les individus se reproduisent de mieux en mieux et exploitent de mieux en mieux les ressources du milieu. Donc si la reproduction est meilleure, l'espèce (ensemble des individus) a moins de chances de s'éteindre.

La compétition est partout : compétition pour l'accès aux ressources (croissance orientée des racines pour accéder aux nutriments), compétition pour la reproduction…C'est l'hypothèse de la Reine Rouge. La dispersion permet l'évitement de la compétition entre apparentés : en cause des émigrations depuis des sites à durée limitée de vie ou la fondation de nouvelles populations.

Par exemple pour les plantes, la dispersion permet de disperser le génotype et d'entrer en compétition avec les autres plantes pour remplacer leur génotype. En effet, il est inutile d'être en compétition avec soi même, d'où la dispersion.

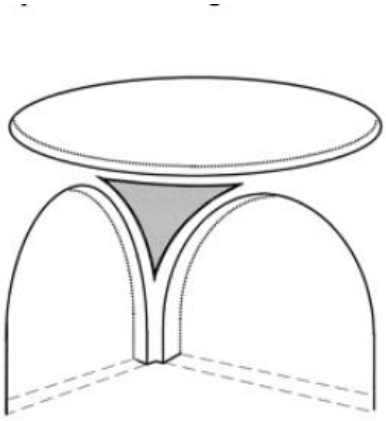
iii. Hasard, usure et contingence

La sélection naturelle n'est pas l'unique processus évolutif. Tout en biologie ne comprend pas à la lumière de la sélection naturelle, il faut tenir compte du hasard.

Les processus d'échantillonnage des gènes (dérive génétique) amènent à perdre la diversité. Les mutation délétères abîment continuellement le pool génétique des populations. Les populations incapables de purger leurs mutations délétères risquent de les exprimer trop fortement voire même

de les fixer (dépression de consanguinité (réduction des critères de sélectivité suite à la reproduction entre individus apparentés) ou fonte mutationnelle (dans les grandes populations, les mutations désavantageuses qui se produisent en permanence et se transmettent sont éliminées par le processus de sélection naturelle. Dans les petites populations, ce processus de sélection n'est plus aussi efficace et aboutit à l'accumulation de mutations délétères, qui peu à peu, vont réduire la capacité de la population à croitre).

Le taux de substitution est le taux de mutations qui restent après sélection. La très grande majorité des mutations sont soit neutres ou délétères. Les mutations avantageuses sont très rares.



En 1979, les biologistes évolutionnistes Richard Lewontin et Stephen J. Gould ont utilisé le terme d'architecture « spandrel » pour illustrer la distinction entre l'évolution d'une caractéristique phénotypique qui a conféré un avantage sélectif par son utilité directe, et celui qui provient simplement comme un produit dérivé de ce processus. Par conséquent l'expression « spandrel » se réfère à l'utilisation adapatée d'une fonction sélectionnée dans un autre but. Une « spandrel » peut ensuite être cooptés pour un autre usage très fructueux selon un processus connu sous le nom « d'exaptation ».. Un pendentif (ou spandrel tridimensionnel) est formé comme un espace forcément triangulaire où un dôme rond réunit deux arcs en plein cintre à angle droit.

II. Code barre moléculaire du vivant

Plusieurs expressions : code barre moléculaire, barcode moléculaire, barcoding moléculaire, DNA barcoding, code-barre génétique.

C'est une méthode d'identification moléculaire fondée sur la comparaison de séquences précises d'ADN présentes dans un échantillon ou un spécimen, vivant ou mort.

Le barcode sert à **aider la taxonomie** lorsque les critères anatomo-morphologiques sont peu distinctifs des espèces. De plus il existe une « crise » de la taxonomie, car peu il y a peu de taxonomistes et cette méthode est également utile pour la « taxonomie traditionnelle ». Le barcode sert à mettre en évidence de nouvelles espèces dans un souci de meilleure connaissance de la biodiversité et de sa conservation. Au taux actuel d'espèces décrite par an et en fonction du nombre d'espèces supposées, il y faudrait entre 1500 et 15 000 ans pour inventorier cette biodiversité.

C'est une méthode standardisée pour fournir un outil supplémentaire pour les taxonomistes. Il ne nécessite pas de connaissances approfondies du taxon (information indépendante du taxonomiste). C'est une méthode relativement rapide pour répondre à la demande d'expertise en biologie de la conservation, agriculture, santé publique, environnement, agronomie. On assigne un individu quelque soit son stade de développement, son état de conservation et le matériel disponible.

Cependant les taxonomistes sont aussi nécessaires dans les différents milieux : dans la biologie de la conservation, on a besoin d'eux pour reconnaître les espèces en danger et lutter contre les espèces invasives. Dans l'agriculture on a besoin d'identifier les espèces nuisibles. Dans la santé publique on a besoin de lutter contre les vecteurs des espèces pathogènes et les pathogènes eux-mêmes. En environnement on a besoin de suivre des espèces indicatrices de la qualité de l'environnement et en agronomie on a besoin de vérifier l'identité des espèces exploitées.

Certains échantillons (larves, oeufs, fragments) ne sont pas reconnaissables via des caractéristiques morphologiques. Très souvent, les échantillons qui sont les pus intéressants à identifier d'un point de vue appliqué sont justement dans ce cas là (saisies en douanes, répression des fraudes). L'identification moléculaire est dans ces cas la seule alternative.

➔Quels outils pour le barcoding moléculaire ?

Un **marqueur génétique approprié** pour le barcoding moléculaire doit répondre à un certain nombre de critères :

- Dans le groupe d'étude, le marqueur doit être suffisamment **variable pour discriminer** entre la plupart des espèces, mais suffisamment **conservée** pour être moins variable au sein d'une espèce.
- Les sites d'amorçage doivent être suffisamment conservés pour permettre une **amplification fiable** sans risque de faux négatifs lorsque l'objectif est l'analyse des échantillons groupés, par exemple, lorsque l'on veut analyser l'ensemble des animaux présent dans un échantillon de sol sans séparer les individus ou analyser de l'ADN de l'environnement telles que l'ADN subfossile résiduel du sol.
- Le gène doit apporter une information phylogénétique suffisante pour affecter les espèces de grands taxons utilisant des approches phénétiques simples.
- L'amplification et le séquençage doivent être aussi robuste que possible, même dans des conditions et des protocoles de laboratoire variables.
- L'alignement des séquences doit être possible également parmi les taxons lointainement apparentés.

Pour les animaux, le choix s'est porté sur une séquence d'environ 650 paires de bases du gène codant **la sous unité 1 de la cytochrome c oxydase mitochondriale (COI).** C'est une enzyme située dans la membrane mitochondriale interne chea les eucaryotes aérobies (contenant du dioxygène). Avec un taux de 2% par million d'années, deux séquences de 600 paires de bases qui ont évoluées indépendamment depuis un million d'années auront donc en moyenne 12 nucléotides de différence. Il y a donc de grands intérêts techniques : amorces universelles, beaucoup de copies par cellule, gène haploïde (non-recombinant).

La séquence du génome mitochondrial permet l'obtention de nombreuses copies dans une même cellule (100 à 10 000 copies) et une variabilité intra spécifique faible (perte du polymorphisme en relation avec l'hérédité maternelle du génome mitochondrial) ce qui rend le génome mitochondrial 5 à 10 fois plus « spécifique » que le génome nucléaire.

Les deux propriétés du génome mitochondrial (faible variabilité intra spécifique et forte variabilité inter spécifique) sont à l'origine d'une véritable frontière génétique bien caractérisée entre la plupart des espèces. **C'est cette propriété qui donne au génome mitochondrial la capacité d'identifier les espèces.**

Enfin, il n'y a pas d'introns dans le génome mitochondrial, ce qui rend plus facile l'amplification des séquences.
Le génome mitochondrial humain est composé de 13 gènes codant des protéines, 2 gènes codant des ARN ribosomaux et 22 gènes codant des ARN de transfert.

➔ Quelles sont les limites du barcoding moléculaire ?

Dans les groupes où la taxonomie est bien connue (les oiseaux par exemple), on peut identifier un spécimen à partir d'une séquence d'ADN (expertise taxonomique). Dans les autres groupes, il reste à

consolider la taxonomie avec tous les outils dont disposent les scientifiques (anatomies, morphologies, molécules…) avant de pouvoir se contenter de n'utiliser qu'un seul gène pour attacher un spécimen à un nom d'espèce.
Le procédé de barcoding se révèle également inefficace dans le cas d'espèces récemment séparées mais aussi dans le cas d'insertion ponctuelle d'un fragment d'ADN dans le génome d'une autre espèce (phénomène d'introgression qui se réalise à l'occasion d'hybridations entre des individus de deux espèces différentes). Les pseudogènes nucléaires (copies de gènes non exprimées) peuvent également entraîner des complications dans l'identification des espèces.
Enfin, cette technique s'applique à tout spécimen ou échantillon tant que de l'ADN en relativement bon état est présent, même en faibles quantités, y compris des échantillons séchés, cuits, broyés et mélangés…Cependant, même si elle peut s'appliquer également à des spécimens anciens, il est impossible de tirer des résultats d'échantillons à ADN trop dénaturé.

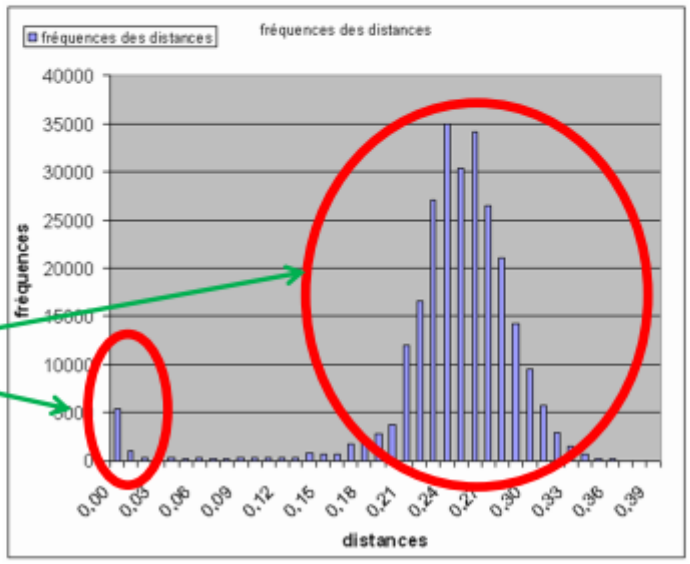
➔ Comment mettre en œuvre le barcoding moléculaire ?

Il faut d'abord une librairie de séquences suffisamment importante pour pouvoir mettre en évidence la frontière génétique entre les variations intra et interspécifiques.

On séquence d'abord un grand nombre d'individus. On calcule ensuite les distances entre chaque séquences à l'aide d'une matrice de distance comme MEGA (Molecular Evolutionary Genetics Analysis). On compare ensuite les variabilités intra et interspécifique. L'utilisation d'une base de donnée de séquences biologiques et l'alignement multiple des séquences permet de détermner un positionnement phylogénétique.

Détermination de la frontière génétique
L'exemple des Mammifères
Base de données : 719 spécimens
L'analyse de la matrice de distance fait apparaître une frontière génétique nette entre les variations inter et intra spécifiques. Ce qui permet d'envisager d'utiliser la séquence COI dans le groupe des Mammifères pour réaliser des caractérisations d'espèces.

Il existe une différence marquée entre les variations intra et interspécifiques



75

L'histogramme ci-dessus présente deux groupes d'individus : Un premier groupe qui présentent entre leurs séquences COI des distances élevées comprises entre 0,15 et 0,36 qui correspondent à des distances entre des individus d'espèces différentes. Un deuxième groupe d'individus qui

présentent entre leurs séquences COI des distances faibles comprises entre 0 et 0,01 qui correspondent à des distances entre des individus de même espèces.

Entre ces deux « populations », il existe des distances intermédiaires mais représentant moins de 1% du total des distances mesurées.

Ces valeurs intermédiaires peuvent provenir de :

- 2 espèces taxonomiquement différentes qui n'en sont en réalité qu'une seule ce qui se traduit par des distances associées à des variations interspécifiques mais qui en fait correspondent à des variations intraspécifiques.
- 1 espèce qui en contient en réalité 2 ce qui se traduit par des distances associées à des variations intraspécifiques mais qui correspondent à des variations interspécifiques.
- Des problèmes d'échantillonnage.
- Des problèmes liés à la qualité de la séquence COI utilisée.

La réalisation de phylogénies avec des séquences de spécimens variés permet de valider l'utilisation d'une séquence de référence pour un taxon donné.

http://www.boldsystemps.org/ est le site qui permet d'explorer les données barcode et d'identifier les séquences (BOLD : Barcode of Life Data Systems). En 2018, on compte plus de 6M de séquences soit 2 fois plus qu'il y a 4 ans.

Cours 7 : Génétique et éthique

I. Les maladies génétiques

Parmi les maladies génétiques on trouve des maladies :

- Monogénique : hémophilies, myopathie de Duchenne, mucoviscidose, maladie de Huntington, certains cancers, immunodéficiences, maladies métaboliques, syndrome de l'X fragile, hémochromatose…Elles sont rares et proviennent de mutations causales.
- Multifactorielles : diabète, infarctus du myocarde, hypertension, asthme, polyarthrite rhumatoïde, obésité, cancers, psychiatrie. Maladies communes avec des facteurs génétiques de prédisposition ou de vulnérabilité ainsi que des facteurs environnementaux.

Dans la génétique on peut parler de traits quantitatifs comme la taille le poids et l'hypertension artérielle. CE sont des traits avec une distribution normale dans la population. On parle d'anormalité pour 2 d.s. au-dessous ou au-dessus de la moyenne. La taille est un trait quantitatif à variabilité continue : 80% de la taille est déterminée par l'hérédité et des variations entre populations et dans le temps font évoluer la taille.

Les traits qualitatifs sont des traits comme l'infarctus du myocarde, la polyarthrite rhumatoïde, la maladie de Crohn, l'asthme, l'obésité, la maladie d'Alzheimer, le diabète, la schizophrénie. Les caractéristiques de traits qualitatifs sont un effet de seuil (seuil dans la population en fonction de la prédisposition. Une fois ce seuil atteint, le trait qualitatif se manifeste. On distingue aussi nettement un phénotype normal et anormal. Il y a une variation continue sous-jacente prédisposant à la malformation avec un seuil conduisant à un phénotype anormal.

A) Les maladies monogéniques

Les maladies **monogéniques** sont liées à la mutation d'un gène et on parle d'hérédité mendélienne (transmission des maladies dues à une mutation dans un seul gène.).

Les lois de Mendel :

- Première loi : Loi de l'uniformité des hybrides de première génération. Lorsque l'on croise deux espèce, la première génération d'hybride est toujours uniforme alors que le génotype et le phénotype des descendants sont hétérozygotes (AA+aa=Aa)
- Deuxième loi : Loi de disjonction des allèles (ou de ségrégation des caractères dans la génération F2): Le croisement d'un hybride par lui-même donne 3 descendants génotypiquement différentes mais phénotypiquement égaux, et 1 génotypiquement et phénotypiquement différent.
- Troisième loi : Loi d'indépendance de la transmission des caractères : On croise la première génération hybride avec un parent récessif (dont les caractères ne peuvent être observés que si ses deux gènes sont égaux)

Les maladies monogéniques représentent environ 5000 maladies. On parle de transmission **autosomique** lorsque les caractères transmis sont portés sur les chromosomes sexuels appelés **autosomes.**
Un parent est dit **sain** s'il ne possède pas le gène et ne risque pas de le transmettre à sa descendance. Un parent est dit **porteur sain** s'il n'est pas atteint par la maladie, mais s'il possède le gène que lui a transmis l'un de ses propres parents. Un parent est dit **atteint** lorsque la maladie s'exprime chez lui.

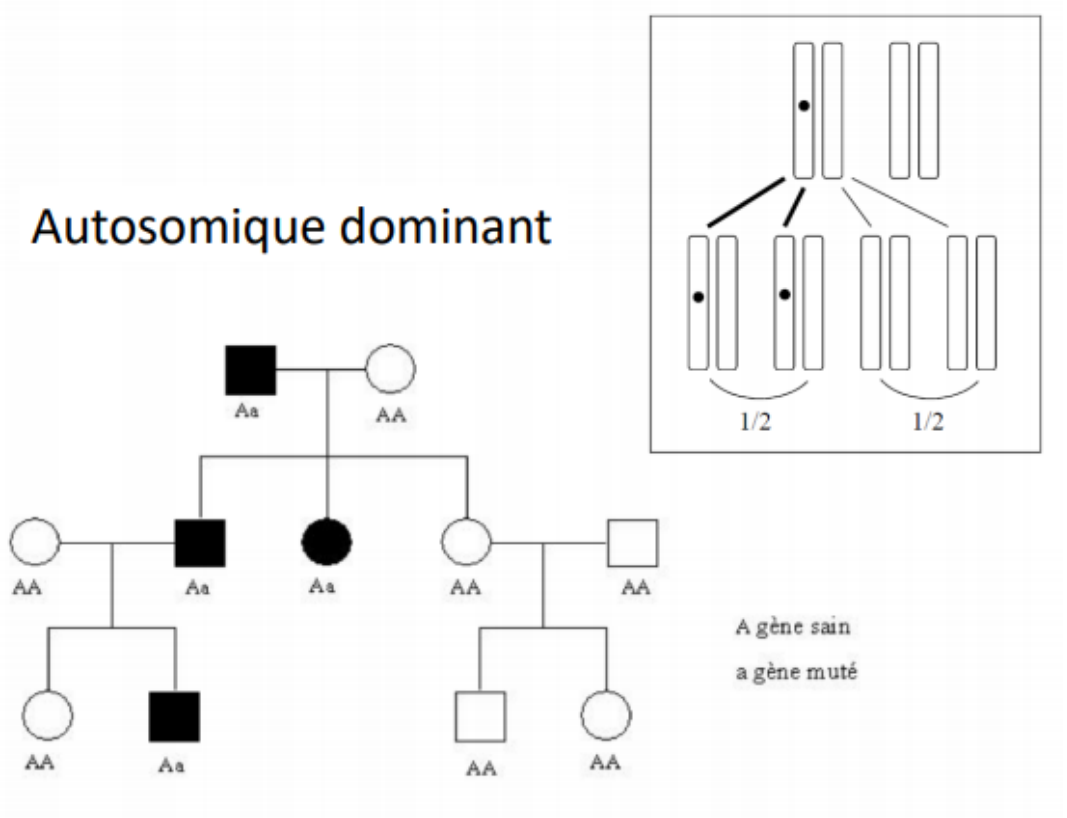
Chaque gène est normalement présent deux fois. Si les deux versions sont **identiques, on est homozygote.** Si les deux versions sont **différentes, on est hétérozygote.** Un gène est **dominant** s'il s'exprime (il permet à lui seul d'avoir une incidence sur le corps.). Ce sont des gènes hétérozygotes avec une copie du gène altéré qui sont altérés. Il est **récessif s'il ne s'exprime pas tout seul.** Ce sont des homozygotes avec deux copies du gène altéré qui sont affectés. Un gène récessif lié à X est un mâle avec une copie du gène altéré sur le chromosome X qui sont affectés.

La transmission **autosomique dominante** est lorsque l'un ou deux parents possèdent le gène dominant et est donc atteint.

Un des parents est atteint et homozygote= 100% des enfants seront atteints.

Un parent est atteint et hétérozygote, l'autre est sain= 50% des enfants risquent d'être atteints.

Les deux parents sont atteints et hétérozygotes= 75% des enfants seront atteints.

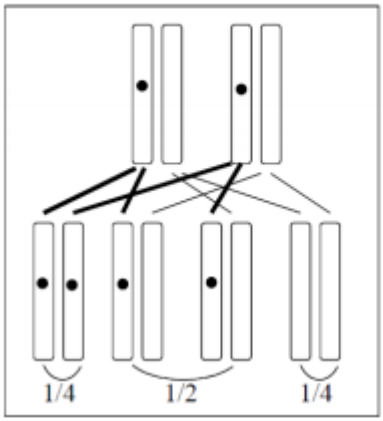


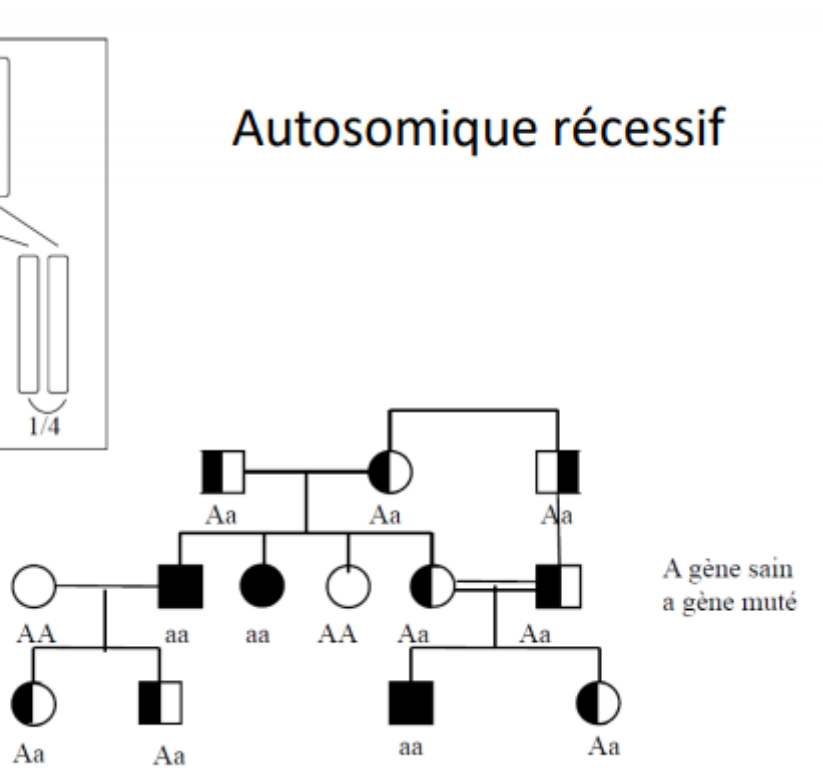
La transmission **autosomique récessive** est lorsque soit l'un des parents présente la maladie, il est atteint et donc porteur de deux gènes déficients ou que l'un des parents possède le gène mais il n'est pas par la maladie.

Dans le premier cas, si l'autre parent est lui aussi atteint, 100% des enfants seront atteint. Si l'autre parent n'est pas atteint mais porteur du gène déficient, 50% des enfants seront atteints et tous seront porteur du gène déficient. Si l'autre parent n'est pas porteur du gène déficient, aucun des enfants ne sera atteint mais tous seront porteurs du gène qu'ils risquent de transmettre à leurs enfants.

Dans le deuxième cas, si l'autre parent est lui aussi porteur sain de la maladie, 25% des enfants risquent de présenter la maladie, 50% seront porteurs sans être atteints et 25% seront sains. Si l'autre parent ne porte pas le gène déficient, aucun des enfants ne présentera la maladie mais 50%

d'entre eux risquent de transmettre la maladie à leur propre descendance. Ce cas fait qu'une maladie héréditaire qui ne s'est pas manifestée depuis des générations peut se manifester brusquement. Ce problème intervient surtout dans des mariages consanguins, intra-ethniques ou intra-régional.





Les maladies monogéniques sont liées au chromosome X.

Certaines maladies d'apparence multifactorielle peuvent être en réalité monogéniques comme l'obésite par mutation du gène LEP. Cette obésite se manifeste de manière sévère, précoce avec hyperphagie mais une croissance staturale normale.

B) Le cas du cancer du sein

Le gène BRCA1 est un grand gène situé sur le chromosome 17q contenant 22 exons codant 1 683 acides aminés. Plus de 500 mutations ou variations de séquence ont déjà été décrites, et le plus souvent une mutation semble unique pour chaque famille.

Près de 90% des familles atteintes d'un cancer du sein ou de l'ovaire de transmission dominante autosomale présentent une anomalie chromosomique de cette région. Pour les familles n'ayant que des cancers du sein, on retrouve l'anomalie dans environ 45% des cas. Certaines familles, n'ayant que des cancers de l'ovaire, ont aussi une anomalie de BRCA1. Les hommes porteurs d'une anomalie BRCA1 ne semble pas avoir un risque accru de cancer.

L'anomalie génétique de BRCA1 est **transmise de façon dominante, autosomale mais avec une pénétrance incomplète.** Environ 50% des enfants portent ce trait. Les femmes porteuses du trait ont un risque de développer un cancer du sein pendant leur vie dans 55% à 85% des cas et un cancer de l'ovaire dans 15 à 45% des cas.

II. Notions de génétique moléculaire des cancers

A) Bases moléculaires de l'oncogenèse.

➔Quelles sont les causes génétiques des cancers ?

Elles sont dues à des altérations génétiques propres au clone tumoral.
- Altérations des oncogènes : gènes dont l'activation (altération monoallélique, mutation gain de fonction) va accélérer la prolifération cellulaire.
- Altération des gènes suppresseurs de tumeur : gènes donc l'inactivation (altération biallélique, mutation perte de fonction) va altérer les mécanismes de contrôle du cycle cellulaire, de l'apoptose (mort cellulaire programmée) ou de la différenciation cellulaire. Ou encore des gènes codant des protéines impliquée dans la détection et la réparation des dommages de l'ADN.

Les protéines codées par les oncogènes : protéines impliquées dans la transduction du signal mitogène, protéines contrôlant le cycle cellulaire, protéines impliquées dans l'apoptose, protéine régulatrice de l'immortalité cellulaire, protéines impliquées dans les processus de métastase (une métastase désigne un foyer de cellules cancéreuses se développant sur un organe mais provenant du cancer affectant un autre organe), protéines pro-angiogéniques (croissance de nouveaux vaisseaux sanguins à partir de vaisseaux préexistants).

Les gènes suppresseurs de tumeurs sont altérés tel qu'au niveau du gène, l'inactivation du gène est une condition essentielle au développement tumoral. Elle se fait soit par des mutations perte de fonction qui sont associées à une perte d'hétérozygotie (ou inactivation due à un mécanisme épigénétique comme la méthylation du promoteur), soit à un syndrome de prédisposition héréditaire aux tumeurs (un patient atteint de mutation constitutionnelle est prédisposé à développer une tumeur, *a contrario*, mutations somatiques dans les tumeurs sporadiques). Ces gènes produisent une protéine capaple d'inhiber la croissance de cellules tumorales *in vitro* par réversion du phénotype tumoral.

Les gènes codant des protéines impliquées dans la détection et la réparation des dommages de l'ADN sont le BRCA1 BRCA2, MLH1, MSH2, PMS1, PMS2.

B) Agrégation familiales de tumeurs fréquentes

a. Cancer du sein

Le cancer du sein dans la population résulte du mélange de deux distributions : l'une sans déterminant héréditaire (tumeurs sporadiques sans composante familiale) (95%), l'autre en rapport avec un allèle de prédisposition, autosomique dominant, de faible fréquence, à pénétrance élevée (5%). Les gènes BRCA1 et 2 ont été clonés en 1994 et 1995.

Le gène BRCA 1 et BRCA 2 provoquent des mutations inactivatrices à l'état hétérozygote sur le gène suppresseur de tumeur. Il peut y avoir des mutations somatiques exceptionnelles dans le as de tumeurs sporadiques.

III. Génétique et éthique

a. Cadre de définition

La génétique questionne l'éthique : de part les applications croissantes en génétique humaine et médiale, la question de la société et de la manipulation génétique mais aussi la religion, la question de la vie et de sa manipulation.

L'intervention sur le patrimoine génétique se base sur le séquençage et le déchiffrement du génome humain pour peut être ensuite offrir des thérapies géniques et une médecine prédictive avec peut être un effet anxiogène mais aussi une future discrimination par les employeurs et assureurs. Il y a aussi un risque d'**eugénisme** (attitude philosophie ou théorie qui préconise une amélioration du patrimoine génétique de certaines populations humaines par la sélection, par l'interruption de la grossesse, par l'interdiction de la reproduction des individus considérés comme inférieurs ou même par élémination).

L'éthique est la partie de la philosophie qui envisage les fondements de la morale. C'est l'ensemble des principes moraux qui sont à la base de la conduite de quelqu'un. La démarche éthique se distingue de la démarche scientifique qui elle se base sur des jugements de fait formulés dans des énoncés descriptifs.

➔Quelle différence entre morale et éthique ?

Mores signifie « mœurs » en latin. Éhtos signifie « mœurs » et *Êthos* signifie « manière d'être » en grec. Cette distinction est à l'origine de la distinction entre morale et éthique. La **morale** est basée sur des normes et des principes généraux, universalisables, variables en fonction des cultures. Elle détermine le bien et le mal. L'**éthique** est nécessaire face à une situation médicale problématique singulière lorsque ni la loi, ni la déontologie, ni la science ne peuvent proposer de réponses simples. Elle est une réflexion, à plusieurs, recherchant le moindre mal pour le sujet malade.

L'éthique recherche le bon, le bien, le juste, le souhaitable, le moins pire. Elle se pose la question de que faire, pour bien faire, comment et pourquoi ? Elle agit par recommandation et obéit à un impératif hypothétique c'est-à-dire selon les circonstances.
La **morale** détermine le bien et le mal et se pose la question de que dois-je faire ? Elle agit par obligation personnelle et collective. Elle obéit à un impératif catégorique et s'applique nécessairement et de manière universelle le plus souvent.

Le **droit** détermine le légal et l'illégal et se pose les questions de savoir comment éviter et juger les conflits et sociétés ainsi que de savoir comment protéger le faible du pouvoir. Il agit par contrainte, protection et sanction. Il s'applique de manière homogène sur tous les citoyens et offre une possibilité de jugement singulier par les juges.

Il est nécessaire de faire des distinctions parmi les interactions entre ces termes : trahir son ami est égal mais pas éthique, déroger à la loi est parfois éthique (lois du parti national-socialiste). Le droit peut engendrer des comportements éthiques. La loi peut reposer sur des repères éthiques (GPA).

L'**éthique** permet de réfléchir à nos actions de façon responsable, aux principes, aux valeurs qui les fondent et aux résultats espérés. Les émotions et les vertus participent de la réflexion éthique qui est une réflexion intégrative des différentes morales. Elle permet de :

- Arracher à l'immédiateté de l'appréciation des situations
- Amener à prendre du recul par rapport à une affectivité plus ou moins impulsive
- Attirer l'attention sur les conséquences proches ou lointaines de ses actes

- Proposer une hiérarchie des valeurs

L'éthique se décline de la réflexion sur les concepts fondamentaux à la démarche pratique.

b. L'éthique de la médecine

L'éthique de la médecine a pour thèmes, selon *Les fondements philosophiques de l'éthique médicale* par Suzanne Rameix, la procréatique (technoscience du début de la vie), l'expérimentation sur l'homme (recherche clinique), l'intervention sur le vieillir et le mourir (soins palliatifs et législation sur la fin de la vie), les greffes d'organes et l'utilisation de parties du corps humain, le cerveau et les manipulations de la personnalité, les interventions sur le patrimoine génétique.

Par exemple, le séquençage des génomes pose des questions éthiques : certes on peut identifier toutes les mutations et gènes de prédisposition mais cela pose un problème de diagnostic prédictif avec un séquençage systématique des nouveau-nés et un bousculement des normes de bonne pratique de génétique internationales. Que faire une fois les informations connues ? Y'a-t-il un risque d'eugénisme ?

c. La génétique médicale

La génétique médicale constitue en le diagnostic clinico-biologique autour d'une maladie génétique mais aussi de nombreux aspects faisant intervenir l'éthique : prise en charge du patient, soutien familial, conseil génétique, dépistage familial, diagnostic prénatal, diagnostic pré symptomatique. C'est la discipline concernée par les maladies génétiques, en particulier dans leur dimension familiale.

Quelques articles de loi pour le dépistage familial

« Lorsque l'examen requiert **d'étudier les caractéristiques génétiques d'un ou plusieurs membres de la famille**, il appartient à la personne concernée, sur les conseils du médecin prescripteur**, d'obtenir le consentement de chacun d'eux**. Ce consentement est donné par **écrit** »

Code de la santé publique, Article R1131-5

« Nul ne peut faire l'objet de **discriminations en raison de raison de ses caractéristiques génétiques** »

Code Civil, Article 16-13

« Les entreprises et organismes qui proposent une **garantie des risques d'invalidité ou de décès ne doivent pas tenir compte des résultats de l'examen des caractéristiques génétiques d'une personne demandant à bénéficier de cette garantie**, même si ceux-ci leur sont transmis par la personne concernée ou avec son accord. En outre, ils ne peuvent poser **aucune question** relative aux tests génétiques et à leurs résultats, ni demander à une personne **de se soumettre à des tests génétiques** avant que ne soit conclu le contrat et pendant toute la durée de celui-ci. »

Code de la Santé Publique, Article 1141-1

« En cas de diagnostic d'une **anomalie génétique** grave, le médecin informe la personne ou son représentant légal des risques que **son silence ferait courir aux membres de sa famille potentiellement concernés** dès lors que des mesures de prévention ou de soins peuvent être proposées à ceux-ci. Toutefois, si elle estime ne pas être en mesure d'informer les membres de sa famille, elle peut demander par un document écrit au médecin prescripteur de procéder à la transmission. Le médecin porte alors à la connaissance des membres de la famille dont il a reçu les

coordonnées l'existence d'une **information médicale à caractère familial** et les invite à se rendre à une consultation de génétique, sans dévoiler le nom de la personne ayant fait l'objet de l'examen, ni l'anomalie génétique, ni le risque qui lui est associé. »

Code de la Santé Publique, Article L. 1131-1-1

3 possibilités pour la personne diagnostiquée : elle informe directement sa famille, elle l'informe indirectement via la procédure de l'information médicale à caractère familial de manière anonymisée ou elle refuse d'informer sa famille. Ce refus sera noté et la responsabilité civile de la personne peut être engagée en cas de procès.

L'intérêt du diagnostic génétique est d'établir un pronostic évolutif ainsi qu'une prise en charge et un suivi spécifique, une surveillance du risque des complications et l'apparition de formes syndromiques liées à des anomalies associées. Le diagnostic permet de produire un conseil génétique avec un diagnostic prénatal/préimplantatoire et un diagnostic familial.

Il existe des tests génétiques en accès libres, des autotests via internet mais ceux-ci sont caractérisés par l'absence d'encadrement médical (indication, interprétation), l'absence d'encadrement éthique et juridique (identité, consentement, mineur) et l'absence d'encadrement technique (fiabilité, qualité). Des exemples de tests génétiques en accès libre : des tests pour les maladies multifactorielles, des tests de paternité, filiation, appartenance à un groupe ethnique, identification des personnes, sexe fœtal, caractéristiques physiques…etc. Des noms comme Navigenic, deCODEme, Human Molecular Life Science ou encore Lab Genetics ressortent souvent.

Le diagnostic présymptomatique est en question avec celle de l'étique.
L'étude du génome par la biologie moléculaire permet de connaître le statut d'un individu vis-à-vis d'une maladie encore non déclarée. Sont concernées les personnes à risque (famille comportant des sujets atteints, majeurs, asymptomatiques, faisant une démarche volontaire.
2 situations de tests : pour des affections pouvant bénéficier d'une prise en charge préventive et/ou curative (cancers) et des affections sans prise en charge préventive et/ou curative (maladies neurodégénératives.

Le diagnostic pré-symptomatique est une activité régie par le décret 2000-570 du 23 juin 2000. Il fixe les conditions de prescription et de réalisation des examens des caractéristiques génétiques d'une personne. Les DSP doivent être effectuées par des équipes pluridisciplinaires déclarées au ministère réunissant des compétences cliniques et génétiques avec un certain type de prise en charge qui suit une déclaration au ministère de la santé. Le DSP est interdit chez le mineur sauf raison exceptionnelle. Il ne doit pas être fait dans des situations d'urgence. Le DSP se fait avec l'accompagnement d'un proche.

Les inconvénients du DPS sont : des conséquences psychologiques (augmentation de l'anxiété, sentiment de culpabilité, altération de l'image de soi, modification de l'attitude parentale), des conséquences sociales (discrimination à l'assurance, au travail et au projet de vie) et le fait de trouver des maladies à expressivité variable.

Les avantages sont en cas de test négatif de lever une incertitude, diminuer l'angoisse et l'anxiété, de faire une action de prévention et de ne pas subir de discrimination.

Le test génétique chez un mineur doit être fait sur la demande des parents car le sujet n'est pas assez informé, on considère qu'il n'est pas consentant et le test ne vient pas de son libre arbitre. Sur des mineurs symptomatiques, on effectue des tests génétiques à visée diagnostique : fraX, mucoviscidose, myopathie, DMD, Steinert, SMA, CMT, Huntington…On a alors un bénéfice diagnostique. Sur un mineur asymptomatique on réalise des tests pour évaluer le risque la probabilité de l'expression d'une maladie génétique connue dans la famille et sur un enfant à risque ou une mutation familiale identifiée. Les tests à ne pas effectuer chez un mineur asymptomatique sont les maladies génétiques graves de révélation tardive : maladies neurodégénératives, prédispositions héréditaires au cancer, test de recherche de sujet porteur caractère récessif, test d'hérédité multigénique ou poly factorielle.

Chez l'enfant, le bénéfice médical individuel de tests génétiques permet la prévention et la prise en charge adaptée pour chaque maladie et chaque enfant. Le test négatif permet d'éviter des investigations de surveillance et le test positif permet une surveillance adaptée et un traitement préventif.

Quelques questions autour de la réalisation d'un test génétique chez le mineur :

D'un point de vue éthique, il y a une privation du droit de ne pas savoir et un non-respect de la confidentialité. D'un point de vue psychologique il y a une perturbation de la relation parents/enfant et une anxiété autour de l'anticipation de la maladie. D'un point de vue légal, l'intérêt de l'enfant doit primer sur celui des parents.

d. La procréatique et le diagnostic prénatal (DPN)/ Diagnostic pré-implantatoire (DPI)
    i. DPN, DPI, eugénisme et trisomie 21

La procréatique interroge le statut de l'embryon qui n'est « ni chose ni personne » ou « personne potentielle ». Elle interroge le statut du corps et sa disponibilité : corps machine chez Descartes/corps propre ou corps vécu/Chair, monisme des phénoménologues. C'est une discipline où il y a un risque de discrimination et d'eugénisme : DPN avec sélection des embryons non souhaité. C'est une discipline qui questionne la finalité de la médecine : guérir, soigner ou intervenir pour des convenances personnelles ?

Le diagnostic prénatal est encadré par des lois :

« Lorsque l'interruption de grossesse est envisagée au motif qu'il existe une **forte probabilité** que l'enfant à naître soit atteint d'une **affection d'une particulière gravité** reconnue comme **incurable** au moment du diagnostic, l'équipe pluridisciplinaire chargée d'examiner la demande de la femme est celle d'un **centre pluridisciplinaire de diagnostic prénatal**. »

Code de la Santé Publique, Article L2213-1

Il y a des limites au diagnostic prénatal : il est très anxiogène, il y a un risque de fausse couche lié aux prélèvements, il n'apporte pas parfois la solution, il n'y a pas de DPN possible pour toutes les pathologies/anomalies du développement mises en évidence, il est le plus souvent non curatif avec à la fin une interruption médicalisée de grossesse (IMG) et possède une variabilité d'expression et de pénétrance.

Le diagnostic pré-implantatoire est un diagnostic sur une ou deux cellules d'une maladie génétique afin de ne transférer dans l'utérus maternel que les embryons reconnus indemnes. Le DPI évite éventuellement une IMG mais s'expose aux risques et aléas de la FIV (fécondation *in vitro*). Le DPI est indiqué en cas d'antécédents du couple avec de multiples IMG sans enfant sain, en cas de génopathie à hérédité autosomique dominant à expression intrafamiliale très variable ou à révélation tardive ou dans le cas d'une nécessité de recours à la FIV pour le couple et dans le cas d'une indication à DPN.

A noter que les tests prénataux pour maladies communes multifactorielles sont interdits

L'**eugénisme** peut être désigné comme « l'ensemble des méthodes et pratiques visant à améliorer le patrimoine génétique de l'espèce humaine. Il peut être le fruit d'une politique délibérément menée par un Etat et contraire à la dignité humaine.

Les anomalies chromosomiques concernent 0,6% de la population (1 naissance/200), 25% des morts néonatales précoces et 10% des produits de fécondation pour ensuite 70% d'avortement spontané avant la 6ème semaine.

La trisomie 21 correspond à 1/2000 naissances vivantes en France. Elle se caractérise par la présence de 3 allèles sur le chromosome 21. Les symptômes sont une déficiens intellectuelle variable, souvent légère, une hypotonie msuculaire et laxité articulaire quasi constantes, fentes palpébrales en haut et en dehors, épicanthus, nuque plate, visage rond, nez petit, pli palmaire unique bilatéral. Les principales malformations et complications sont : cardiaques et digestive, maladie de de Hirschsprung, petit taille, syndrome de West, épilepsies, leucémies, apnées du sommeil, déficits sensoriels, pathologies auto-immunes et endocriniennes, vieillissement précoce et maladie d'Alzheimer. L'espérance de vie médiane est aujourd'hui supérieur à 50 ans.

ii. DPNI : Dépistage prénatal non invasif

En 1997, on découvre que l'ADN fœtal libre circule dans le sans maternel. Il provient des cellules trophoblastiques (couche cellulaire qui fournit les nutriments à l'œuf et apparaît au 6ème jours après fécondation) en apoptose (mort cellulaire programmée) qui sont relarguées de manière continuelle dans la circulation sanguine maternelle. L'ADN fœtal libre apparaît donc de manière précoce mais se trouve en quantité faible.

On peut alors trouver le génome fœtal entier dans cette circulation. L'ADN fœtal et l'ADN maternel est similaire. On va donc comparer les ADN en mettant en évidence une sur-représentation d'ADN circulant lorsque le fœtus est porteur d'une aneuploïdie (cellule contenant un nombre anormal de chromosomes). Le séquençage à haut débit des séquences fœtales et maternelles de l'ADN circulant va permettre de calculer une perte ou un excès de séquence par rapport à un groupe contrôle euploïde (contraire aneuploïde). La puissance de discrimination augmente avec le nombre de molécules d'ADN séquencées.

e. Le clonage et la modification génétique

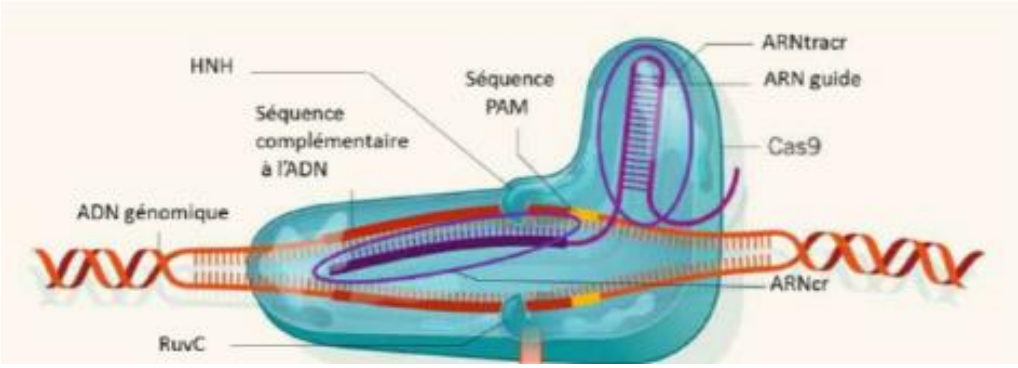
Les animaux d'un clone réalisés par transferts de noyaux provenant d'un même individu, ont le même génotype. Ils sont de même sexe et présentent des caractéristiques physiologiques et morphologiques identiques : caractères sanguins et immunologiques…ou très proches, taille, poids. Les variations constatées entre individus sont très informatives car elles permettent de déterminer si un caractère phénotypique ne dépend que de facteurs génétiques. Par exemple les taches pigmentaires dorsales du pleurodèle semble dépendre de facteurs épigénétique. La taille et poids résultent de l'action des facteurs d'environnement.

Le clonage nécessite l'utilisation de noyaux provenant d'embryons très jeunes ou de noyaux différenciés reprogrammés. Son application aux mammifères fut réalisée en tenant compte de ces impératifs et après adaptation de la technique.

Le 1er clonage réussi chez un mammifère en utilisant des noyaux de cellules d'adultes, fut réalisé chez le mouton. « Dolly » fut obtenue en 1997 par greffe de noyaux de cellules de glandes mammaires. Ces cellules furent mises en culture en milieu dépourvu de sérum : c'est un milieu appauvri. Leurs noyaux passent ainsi hors cyle de division. Après greffe, ils pourront, par cette reprogrammation, se synchroniser avec le cytoplasme des ovocytes dans lesquels ils seront greffés et se comporter comme des noyaux de fécondation. Au total 277 transferts de noyaux furent effectués, 19 blastocystes implantés, 1 embryon viable obtenu qui deviendra la brebis « Dolly ». Cette technique a été réutilisée pour des singes.

Elle a même été utilisée en implantant des fibres de cerveaux malades humains pour que les clones développent les maladies et que nous puissions les étudier. Ainsi on peut constituer des groupes contrôles dont la variabilité génétique est moindre, donc moins d'animaux sont nécessaires. Les cerveaux des primates sont utilisés pour étudier les troubles psychiatriques humains et les maladies neurodégénératives comme Alzheimer et Parkinson.

CRISPR Cas9 est un système de défense des bactéries contre les infections virales. Le système ARN-protéines CRISPR Cas9 reconnaît, ouvre et coupe l'ADN viral et élimine l'ADN viral. En génétique, ce système est utilisé pour modifier le génome.



L'ARN guide (ARNg) comprend l'ARNcr (ARN cripsr) qui cible le locus désiré par complémentarité des bases et l'ARNtracr (trans-activating RNA) qui permet le recrutement de la Cas9. La Cas9 est une nucléase qui effectuera une double coupure trois nucléotides en amont de la séquence PAM (Protospacer Adjacent Motif). Elle effectue ces coupures grâce à deux sous-unités catalytiques indépendantes quji sont HNH du côté de fixation de l'ARN guide et RuvC du côté opposé.

A la suite de l'induction d'une cassure double brin par le système CRISPR-Cas9, il peut y avoir réparation par NHEJ (Non-Homologous End-Joining) s'il n'y a pas de matrice de recombinaison. Cela peut induire le KO du gène par Insertions/délétions.
Si une matrice de réparation est apportée, il va y avoir insertion de la matrice grâce à ses régions d'homologie, par le mécanisme de Recombinaison Homologue.

f. Les expérimentations sur l'homme

Les essais cliniques doivent être soumis pour approbation au Comité de protection des personnes (CPP) depuis la loi Huriet (1988) puis Jardé (2012). Le consentement doit être libre et éclairé. L'essai sans bénéfices doit être justifié. Il y a des expérimentations sur les mineurs et sur les patients protégés (curatelle, tutelle et personnes privées de libertés. Il y a forcément un groupe sous placebo pour un groupe contrôle.

Comment marche le clonage ? On va prélever une cellule totipotente (cellule embryonnaire encore indifférenciée), le noyau va être transféré dans un ovocyte, on va alors avoir une cellule pluripotente (capacité d'une cellule à se différencier en trois feuillets embryonnaires (états embryonnaires) ou cellules germinales) qui va finir par se différencier. La différenciation dépend de facteurs épigénétiques notamment. On parle alors de reprogrammation cellulaire.

En Chine, un chercheur a manipulé le génome d'un embryon pour qu'il ne puisse pas avoir le SIDA plus tard (sachant qu'un des parents l'a eu). Il a inactivé le gène CCR5 qui encore un récepteur que le HIV utilise pour pénétrer les cellules.

Cours 8 : Les faux-semblants

I. Evaluation statistique de la similarité et de la différence

a. Métrologie

La métrologie est la science des mesures, c'est l'ensemble des techniques permettant d'évaluer des grandeurs, des performances

Une **mesure** (résultat obtenu) est le résultat du **mesurage** (action de mesurer, par exemple, doser le taux de glucose) d'un **mesurande** (la grandeur mesurée, par exemple, la glycémie)

La mesure donne un nombre avec une unité. On connaît les incertitudes de la mesure. L'objectif est donc d'obtenir des mesures fiables c'est-à-dire avec un niveau de confiance élevé.

➔Quels sont les critères de validation des méthodes de mesures ?

Fidélité : caractéristique d'une méthode donnant **toujours le même résultat** ou des résultats voisins si elle est **répétée** sur le **même échantillon**

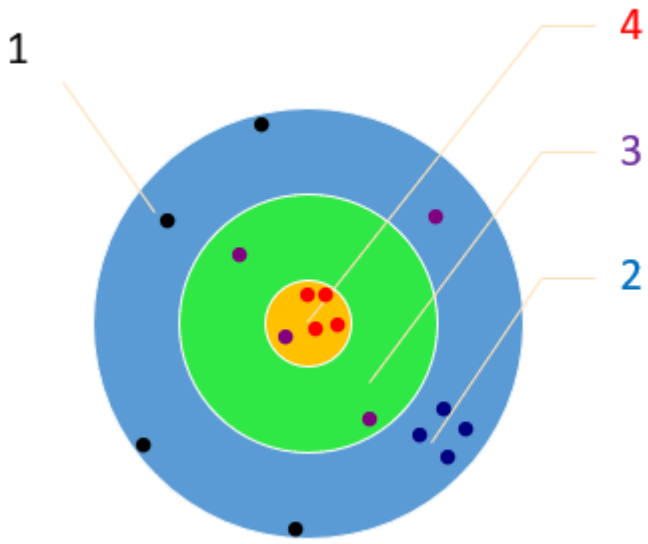
Répétabilité : fidélité sous des conditions de répétabilité (même **méthode, laboratoire, opérateur, équipement** et pendant un **court** intervalle de temps)

Reproductibilité : fidélité sous des conditions de productibilité (même **méthode** dans **différents laboratoires, opérateurs, équipements**) définies selon les textes de références, en admettant la possibilité de conditions de travail différentes.

Justesse : étroitesse d'accord entre le **résultat d'une mesure** et la **valeur attendue** (cible ou valeur réputée vraie.

➔Comment représenter les critères de fidélité et de justesse sous forme schématique ?

Une cible dont le centre représente la valeur attendue de la mesure et dont les tirs représentent les résultats de différentes mesures (en conditions de répétabilité ou de reproductibilité) peut représenter ces critères.



1=Des tirs éloignés du centre et non groupés représente une mesure **non fidèle et non juste**.

2=Des tirs éloignés du centre et groupés représente une mesure **fidèle** mais **non juste.**

3=Des tirs proches du centre et non groupés représente une mesure **non fidèle** mais **juste**.

4=Des titres proches du centre et groupés représente une mesure **fidèle et juste.**

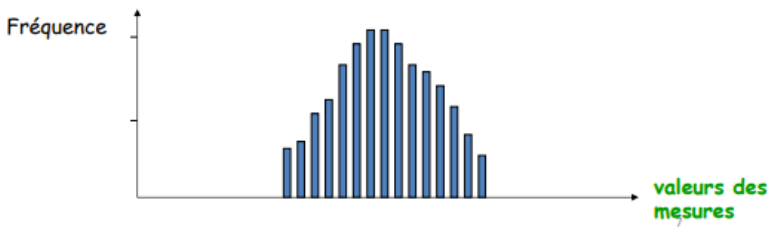
➔Quelles sont les limites d'une mesure ?

La **sensibilité** désigne le rapport entre la variation instrumentale et la variation de la mesurande. Cette sensibilité peut donner des mesures différentes et réduire la fidélité et la justesse.
La **limite de détection** désigne la plus petite quantité ou centration distinguable d'un « blanc » (c'est-à-dire la plus petite quantité mesurable que l'on peut distinguer de rien).
La **limite de quantification** désigne la plus petite quantité ou concentration pouvant être mesurée avec un risque d'erreur connu.

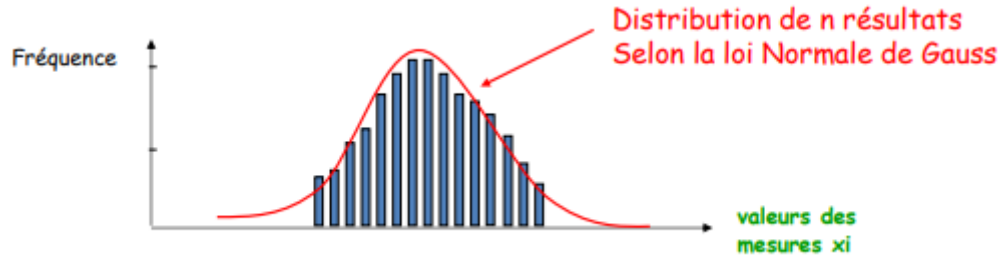
b. Variabilité technique et biologique

On mesure la variabilité technique à partir d'un sujet à partir duquel on fais plusieurs tests et on mesure la variabilité biologique à partir de plusieurs sujets qui subissent un test chacun. Dans le cas de la variabilité technique on cherche à déterminer la justesse et la fidélité d'un instrument technique. Dans le cas de la variabilité biologique on cherche à déterminer la justesse et la fidélité d'une mesure biologique.

Dans le cas de la variabilité technique, la mesure répétée sur un même échantillon donne rarement deux fois exactement le même résultat. Si on trace sur un histogramme la fréquence des résultats en fonction des valeurs obtenues pour classer les différences mesures on obtient en général un modèle en forme de « cloche » que l'on peut représenter ainsi :

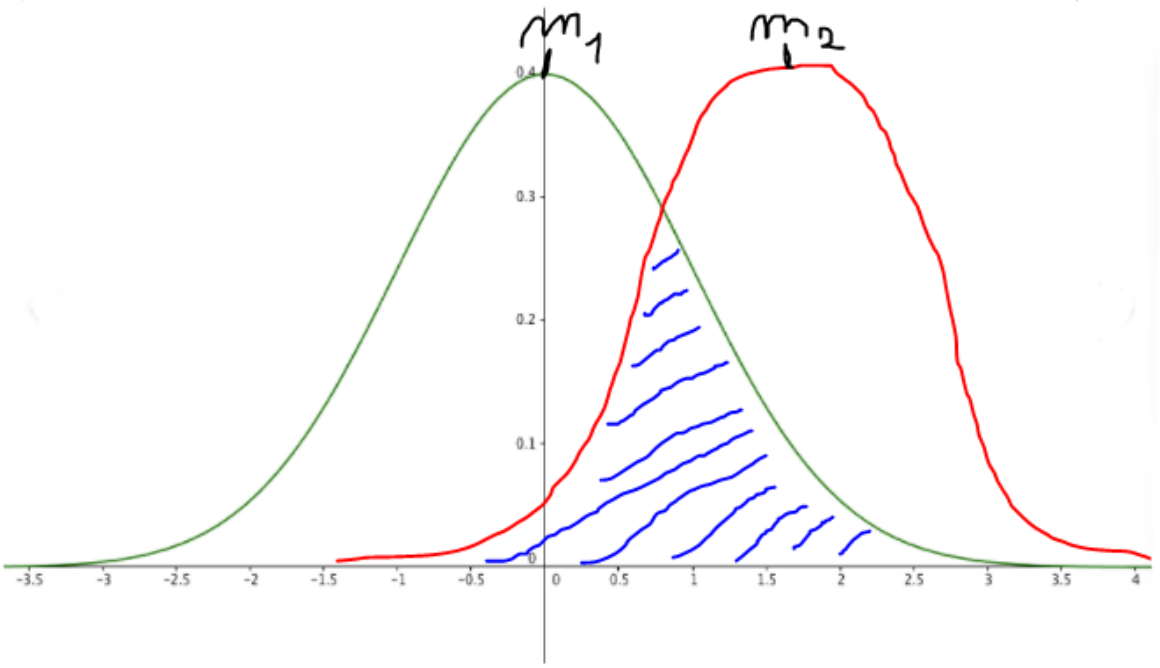


On dit que les mesures suivent le **modèle de Gauss**. Dans un nombre infini de mesures (n) sur le même échantillon biologique et dans les mêmes conditions expérimentales, la série de résultats avec des fréquences variables suit une répartition selon la loi Normale de Gauss.

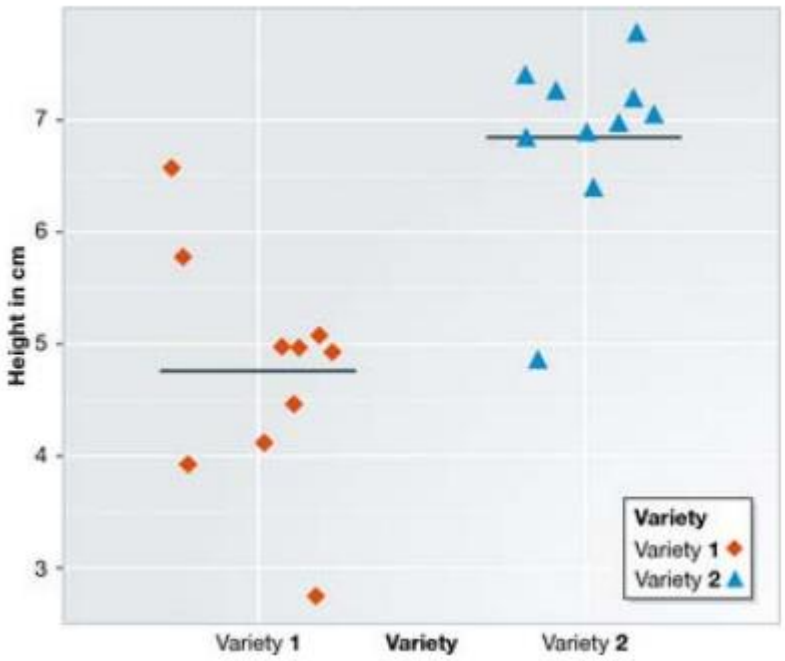


La loi normale de Gauss (dont les caractéristiques seront étudiées au point suivant) est une loi de probabilité dans laquelle les mesures sont des variables aléatoires notées ici x_i pouvant prendre toute une série de valeurs non prévisibles.

Deux courbes qui correspondent à deux séries de mesures de deux échantillons biologiques peuvent donner un recourbement, c'est-à-dire un chevauchement des courbes. En comparant les moyennes (ci-dessous **m_1 et m_2**) on peut conjecturer s'il existe une variation significative en fonction des valeurs.
Par exemple, si on mesure la taille des filles et des garçons d'un groupe et qu'on représente les résultats sous forme de courbes, on va avoir peut-être un recourbement. Dans ce recourbement (que l'on peut représenter **ci-dessous en bleu**) on aura des filles ET des garçons qui font environ la même taille. Dans cet intervalle il n'y a donc pas de variation signification en fonction des valeurs. Si on regarde les moyennes il y a plus de chances qu'il existe une variation significative des valeurs.



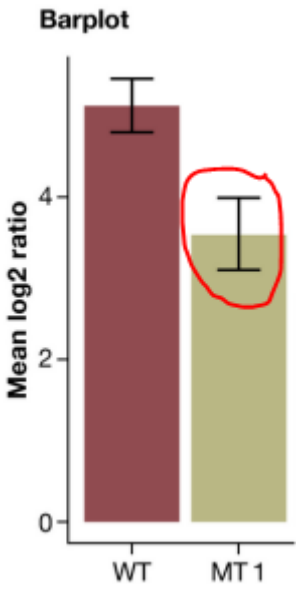
c. Comparaison de moyennes



En comparant deux groupes de mesures et leurs moyennes sur un graphique de « nuage à points », on peut remarquer une variabilité de résultats différentes. Ci-dessous on peut voir que la mesure 2 a moins de variations que la mesure 1 en orange car la mesure 2 a un groupe plus rapproché de sa moyenne que la mesure 2 qui a un groupe plus dispersé.

Pour mesurer la dispersion de part et d'autre de la moyenne il faut mesurer l'ordre de grandeur des valeurs par rapport à la moyenne.

La variance est la moyenne des **carrés** des **écarts à la moyenne**. Elle est représentée sous la formule simplifiée: $\frac{\sum (x - \overline{x})^2}{n}$ avec Σ qui est le symbole de la somme : on a donc somme des écarts (x) à la moyenne représentée par le symbole $\bar{x}$ divisés par (n) l'échantillon. C'est donc bien la moyenne des carrés des écarts à la moyenne.



L'écart type est la racine carrée de la variance. Elle est représentée sous la formule : $\sqrt{\frac{\sum (x - \overline{x})^2}{n}}$ et sous le symbole sigma **σ.**

L'écart type mesure la **dispersion autour de la moyenne**. L'écart type et la **déviation standard** (ou *standard deviation* en anglais) sont des synonymes.

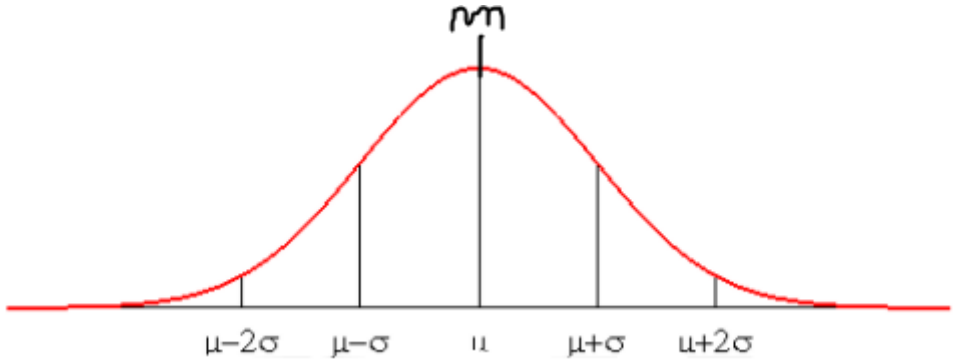
L'erreur standard (ou en anglais *standard error, SE, Standard Error of the Mean, SEM*) est une **estimation de l'écart type lié à l'erreur de la mesure**. L'erreur standard est considérée comme l'écart type de la distribution théorique de toutes les erreurs qui seraient commises au cours des mesures. Cette distribution des erreurs est estimée normale(gaussiene). Si le paramètre ou la statistique est la moyenne on parle **d'erreur standard de la moyenne**. C'est l'écart-type de la distribution d'échantillonnage de la moyenne.

L'erreur standard de la moyenne d'une population est égale **à son écart type divisé par la racine carrée de l'effectif** tel que $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

Lorsque l'on veut décrire la **dispersion** d'une valeur dans une population on utilisera donc préférentiellement l'**écart type ou déviation standard**. Lorsque l'on veut rapporter une mesure que l'on a répété pour **vérifier sa validité**, on utilisera préférentiellement l'**erreur standard**.

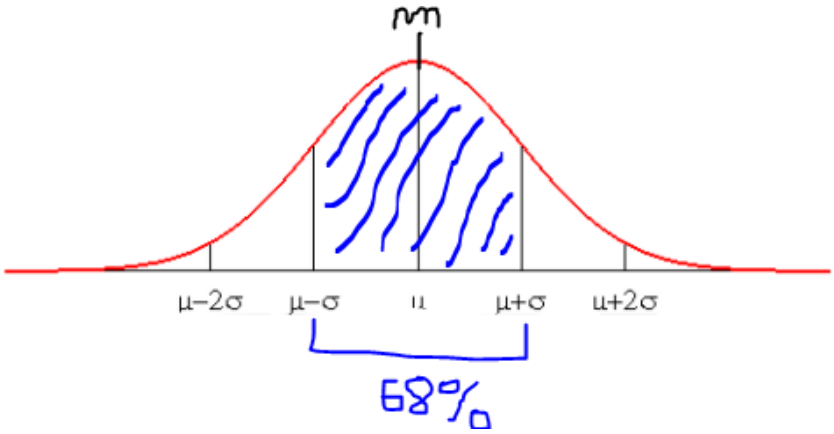
Sur un graphique, on peut représenter l'erreur standard par des intervalles de confiances à 95% qui estiment l'erreur de la mesure (voir graphique ci-dessus page précédente et cercle rouge)

➔Quelles sont les caractéristiques d'une courbe de Gauss

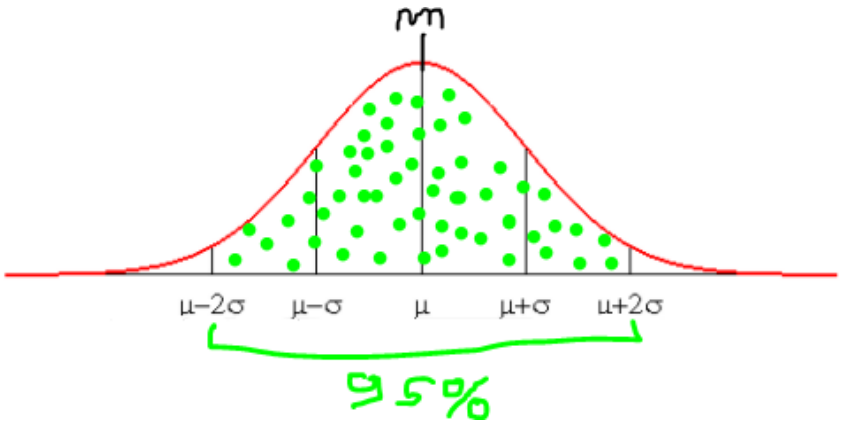


Une courbe de Gauss est une distribution de courbe en cloche. Elle présente les effectifs maximums aux alentours de la moyenne et on observe une décroissance symétriquement de chaque côté. La **moyenne m** partage symétriquement la courbe en cloche. La **médiane** est la valeur qui partage la population en deux sous populations égales et non équivalente à la moyenne arithmétique si la distribution est non gaussienne. La **dispersion s ou écart-type** caractérise la plus ou moins grande dispersion des valeurs autour de la moyenne **m**. C'est la mesure de la distance séparation le point d'inflexion au trait **m**. Plus l'écart type est **grand**, plus la **dispersion est grande**. A noter que la moyenne est aussi notée mu **μ.**

Environ **68%** des mesures sont comprises dans l'intervalle m ± σ c'est-à-dire l'intervalle [μ- σ ; μ+ σ]



Environ **95%** des mesures sont comprises dans l'intervalle m ± 2σ c'est-à-dire l'intervalle [μ- 2σ ; μ+ 2σ]



Environ 99,8% des mesures sont comprises dans l'intervalle m ± 3σ c'est-à-dire l'intervalle [μ- 3σ ; μ+ 3σ]

L'**intervalle de confiance à 95%** est un intervalle de valeurs qui a **95%** de chance de contenir la vraie valeur du paramètre estimé.

d. Intervalle de confiance

L'**intervalle de confiance** mesure le **degré de précision** que l'on a sur les estimations issues de l'échantillon. Il y a deux sources principales de variations sur les données qui peuvent être la cause d'un manque de précision dans l'estimation d'une grandeur :

- **Un nombre insuffisant de données** : dans le cas d'un sondage, on ne sonde qu'une fraction de la population. Dans les mesures physiques, on n'effectue qu'un nombre fini de mesures alors qu'il faudrait souvent en théorie pouvoir en faire une infinité pour obtenir un résultat parfait.
- Il peut également y avoir **du bruit** dans la mesure des données, ce qui est pratiquement toujours le cas pour la mesure des grandeurs physiques. Un **bruit de mesure** est un signal qui vient perturber la mesure : il peut être d'origine naturel (parasites atmosphériques (rayonnements électromagnétiques par exemple), solaires, cosmiques) ou artificiel (parasites industriels).

II. Réflexions autour de la corrélation et du lien de causalité

a. Corrélation et régression

La **corrélation** cherche à mesurer la **force, la rigidité** de la liaison statistique entre X et Y. Si cette liaison est rigide, il sera en particulier possible d'avoir une bonne idée de **Y** en connaissant seulement **X** et vice versa.

Exemple : s'il existe une bonne corrélation entre la taille d'une dent et la taille de son propriétaire chez les tyrannosaures, alors il est possible de déduire de manière approximative la taille d'un spécimen fossile dont on a juste retrouvé une dent et réciproquement.

La régression (linéaire) cherche à caractériser la **pente** de la droite pouvant résumer au mieux la relation entre X et Y une fois choisies des unités de mesures pour X et Y.

Exemple : Si la dose efficace d'un anesthésique est 5mg/kg de poids de corps, on note une pente de 5 pour 1 avec ces unités-là. Un gain de poids de 10kg chez un patient obligera pour le même effet anesthésique à augmenter la dose de 10*5=50mg.

La pense en question n'a d'**intérêt** que si **la relation entre la dose efficace et l'effet** est suffisamment rigide. Si cette relation est en réalité très floue. Le risque de sous-doser ou de sur-doser l'anesthésique devient inquiétant. D'où l'intérêt de **connaître la force de la liaison en plus de sa pente**.

La corrélation et la régression poursuivent le même but : caractériser la liaison statistique entre deux variables quantitatives. **Corrélation et régression** permettent toutes deux de répondre aux questions « Y'a-t-il un lien **statistique** entre X et Y ? » et « Quelle est la **force** de ce lien éventuel (**corrélation**) et quelle est la **pente** de la relation éventuelle pour un jeu d'unités donné (**régression**) ? »

Tableau récapitulatif :

Corrélation	**Régression**
« Y'a-t-il un lien statistique entre X et Y ?	
« S'il existe, quelle est la **rigidité** de ce lien ? En particulier quelle est la fraction de la variance de Y qui subsiste si je **fixe** X ? »	« Quelle est la **relation numérique** entre X et Y pour un jeu d'unités données (si X vaut « telle valeur » dans les unités choisies pour exprimer X, quelle sera **en moyenne** la valeur de Y dans les unités choisies pour exprimer Y ?
Test du coefficient de corrélation (CC)	Test de la pente
➔ Si le coefficient de corrélation est significatif, alors : Il existe **un lien.** Là encore, **impossible** de déterminer la nature du lien sur la seule base du test, sauf dans le cas expérimental où l'on conclurait à un lien causal. La valeur du CC noté R et du coefficient de détermination (CD) noté R^2 ont un sens, et permettre de répondre à la question précédente.	➔Si le test de la pente est significatif, alors : Il existe un **lien,** mais un test significatif est à lui seul **insuffisant** pour démontrer que c'est X qui agit sur Y, sauf dans une situation expérimentale dans laquelle tous les facteurs sont strictement contrôlés et que l'expérimentateur fait varier X. Le calcul de la **l'équation de la droite de régression** est justifié et permet de répondre à la question précédente. Le calcul de CC permet de connaître la force du lien entre X et Y.
➔Si le CC ou le teste de la pente ne sont pas significatif, alors : Soit le lien existe mais il n'y avait **pas suffisamment de données** pour le mettre en évidence. Soit le lien **n'existe pas**.	

La corrélation entre deux variables quantitatives X et Y est l'existence d'une liaison **statistique** entre elles, quelle qu'en soit la raison. « Liaison statistique » signifie ici que les deux variables **ne semblent pas varier indépendamment l'une de l'autre** : connaître la valeur de la variable X ou Y pour un individu fournit une information sur sa valeur pour la variable Y ou X.
L'existence d'une liaison statistique ne signifie **pas nécessairement l'existence d'une liaison causale** (c'est-à-dire un lien de **cause à effet**, une action de X sur Y ou inversement). Si une liaison statistique est décelée entre deux variables, l'existence d'une liaison causale directe entre elles sera seulement **une des possibilités à explorer**. La confusion entre « X et Y sont corrélées significativement » et « X a un effet sur Y (ou inversement) » est probablement **une des erreurs les plus communes** de toute l'analyse des données.

b. Notion de covariance

En statistiques lorsque **deux variables évoluent dans le même sens**, on dit qu'elles **covarient** de manière **positive.**

Exemple : Taille et longévité des mammifères varient globalement dans le même sens, les espèces de grande taille vivent en général plus longtemps que les espèces de petite taille. Mais attention, il n'y a **aucun lien de cause à effet, de causalité** entre taille et longévité : faire plus d'1m80 ne va pas faire vivre plus longtemps

Lorsque **deux variables évoluent dans un sens contraire** l'une par rapport à l'autre, on dit qu'elles **covarient** de manière **négative**.

Exemple : Plus une espèce d'oiseau a une vitesse de croisière **élevée**, plus son rythme cardiaque au repos est **lent**.

Une **covariance négative** peut démontrer une **corrélation**, même si négative. Sur notre exemple précédent, si l'on raisonne à l'échelle de l'espèce, cette relation paradoxale *a priori* s'explique car les espèces qui volent vite sont en général de grande taille alors que les espèces de petites tailles volent assez lentement. Les grands animaux ont des rythmes cardiaques bien plus lents que les petits. Cependant, cela ne montre pas un lien de **causalité** ou de **cause à effet** : la vitesse **en vol** ne peut influencer le rythme cardiaque **au repos** et inversement car ce sont deux états différents.

Lorsque **deux variables sont *a priori* totalement indépendantes** l'une de l'autre on dit que leur **covariance est nulle**.

Exemple : Le lien entre le taux d'accident du Tram B entre Saint Nicolas et Peixotto par jour et la prise d'anti-dépresseurs quotidienne dans une zone englobant les quartiers Saint Nicolas et Peixotto.

c. Notion de coefficient de corrélation

Le **coefficient de corrélation de Pearson** est conçu pour mesurer précisément la liaison statistique entre deux variables qui évoluent **proportionnellement** l'une par rapport à l'autre et qui le font de manière **constante/monotone.** Le coefficient de corrélation inventé par Karl Pearson est désigné par la lettre R et se calcule de cette manière :

$$R = \frac{\mathrm{cov}(X, Y)}{(sX \times sY)}$$

Avec X et Y, deux variables, **cov** la covariance estimée à partir des données et **sX** et **sY** les écarts-types des variables X et Y estimés à partir des données.

Dans sa forme plus « mathématique » :

$$R = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

Tel que la **covariance** se calcule comme la **variance** de **X** multipliée à celle de **Y** (formule vue précédemment).

d. Corrélation, cause, effet, multicausalité

« Corrélation n'est pas raison » est un diction statistique important. Il rappelle que le fait de **trouver une corrélation même élevée et hautement significative entre la deux variables X et Y ne démontre pas du tout qu'il y a un lien de cause à effet entre X et Y et inversement.**

Dans une situation d'observation (c'est-à-dire en dehors d'un dispositif expérimental dans lequel tous les facteurs sont strictement contrôlés), si on constate une forte corrélation entre deux variables, il faut donc impérativement résister à l'envie de proclamer tout de suite un lien de cause à effet.

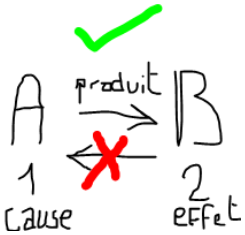
L'établissement d'un lien nécessite une expérimentation en conditions contrôlées ou une accumulation d'autres observations dans **différentes situations**, qui permette d'**exclure** les autres **possibilités non causales**.

Un grand classique est la corrélation observée systématiquement entre deux variables X et Y lorsqu'elles sont toutes les deux **influencées par la même troisième variable Z.**

Exemple : Le nombre Y de coups de soleil attrapés sur une plage est fortement corrélée à la température moyenne X de l'air ce jour-là. X n'influence pas Y. En réalité, la variable Z, la quantité de rayonnement solaire est causale de la température moyenne X de l'air et du nombre Y de coups de soleil attrapés sur une plage.

En toute rigueur, une **corrélation** n'est ainsi qu'un indice statistique précisant le **degré de liaison** entre deux variables. Dès lors que cette relation est solidement démontrée un certain nombre de fois, on peut lui donner **force de loi**.

On est alors tenté d'essayer de la « justifier », de **déterminer ce qui explique le lien entre les phénomènes**. L'intérêt étant de mieux comprendre la relation pour parvenir à mieux prédire l'apparition de l'un des phénomènes en fonction de l'occurrence d'un autre.



On peut définir alors la causalité comme elle est définie en sciences physiques :

« En physique, le principe de **causalité** affirme que si un **phénomène** (nommé **cause**) produit un autre phénomène (nommé **effet**), alors **l'effet ne peut précéder la cause**. L'effet ne peut pas être en théorie la cause de la cause, mais certaines théories envisagent une causalité inversée.

On comprend que la **causalité** est le fait que si un phénomène A produire un phénomène B alors A est la cause de B et A précède B : B n'existe pas avant A.

Lorsqu'on observe une corrélation entre deux phénomènes A et B, il existe au moins six façons d'en rendre compte :

- A est cause de B
- B est cause de A
- A et B ont pour cause un même phénomène-source C
- A est la cause de C qui est la cause de B ou B est la cause de C qui est la cause de A
- A est la cause de B et dans le même temps B est la cause de A (les deux phénomènes se renforcent entre eux)
- La co-occurrence de A et B est une coïncidence

Les cinq premières établissent une **causalité**, c'est-à-dire un rapport de **cause à effet** entre les évènements A et B dans un sens ou l'autre, ou en passant par un intermédiaire C.

La **multi-causalité** est le fait d'avoir de multiples liens qui unissent la cause à l'effet. Comme nous l'avons vu précédemment, il existe des maladies multifactorielles comme l'obésité.

On peut devenir obèse par la génétique, l'absence d'activité physique, le style de vie, le métabolisme basal, l'environnement, les facteurs psychosociaux, la thermogenèse adaptative…Ces liens vont unir la cause (trop d'énergie acquise par rapport à l'énergie dépensée dans la balance énergétique) à l'effet (l'obésité).

Non contents de s'arrêter à l'interprétation erronée d'une corrélation (elle-même déjà fantasmée ou non), certains se permettent d'en déduire des propositions d'action pour remédier à certaines situations jugées non désirables. C'est en effet le stade logique suivant le constat d'existence et la tentative d'explication : la recommandation d'actions ou de comportements.

Le lien existe, on sait d'où il vient, mais qu'en faire ? Le « saut quantique » depuis le stade de la compréhension jusqu'à celui de la prescription est loin d'être anodin, car pas du tout inclus dans les faits eux-mêmes : ici aussi, il n'y a bien qu'un cerveau humain pour imaginer comment passer de l'un à l'autre. N'oublions pas que l'Histoire est remplie d'actes issus de processus cognitifs relatifs à l'effet Cigogne (confondre corrélation et causalité).

III. Homologie et analogie. La convergence évolutive aux différents niveaux d'intégration du vivant

a. Notion d'homologie

« En biologie, l'**homologie** désigne un caractère propre à plusieurs espèces hérité d'un ancêtre commun. Les structures en question qu'elles soient d'ordre anatomique, moléculaire ou génétique, partagent donc une histoire évolutive »
Les chauves-souris et les hommes partagent des bras, ce sont des caractères homologues.
Inversement, deux structures peuvent se ressembler en forme et fonction mais peuvent être apparues indépendamment dans plusieurs groupes différents. Par exemple : les ailes des oiseaux et des insectes.

b. Notion d'analogie

L'exemple cité précédemment est un exemple de caractères analogues.
« Une **analogie** est une ressemblance fonctionnelle entre des traits anatomiques chez différentes espèces. Ils ont généralement la même fonction biologique, mais en revanche pas obligatoirement la même origine évolutive. Ils ne partagent pas d'ancêtre communs. »

c. Notion de convergence évolutive

« En biologie de l'évolution, la **convergence ou homoplasie** est la présence chez deux espèces de caractères analogues, d'une même adaptation, mais qui n'a pas été hérité d'un ancêtre commun. Elle résulte de deux évolutions indépendantes dans un même type d'environnement »
Par exemple : les cétacés font parti de la classe des mammifères même s'ils ressemblent fortement aux poissons. Ils ont des évolutions convergentes ayant mené dans les deux cas à une adaptation au milieu aquatique.

Pour ne pas confondre : Un caractère est **homologue** quand deux espèces ont hérité de ce caractère d'un ancêtre commun. Un caractère est **homoplastique** quand deux espèces provenant d'un même ancêtre commun évoluent différemment mais possèdent ce même caractère ne provenant pas d'un même ancêtre commun.

d. Notion de réversion

« La **réversion** est, en génétique, plus précisément dans le domaine de la biologie de l'évolution, le **retour** d'un caractère à un **état primitif** se manifestant par une mutation génique dans l'ADN permettant de **retrouver** les fonctions d'un gène après qu'il les a perdues à la suite d'une première mutation. […] Il existe deux types de réversion :

- Les réversions vraies : le retour à l'état initial du gène
- Les réversions fausses : l'état est différent de l'initial mais permet le rétablissement de la fonction »

La réversion biologique remet en cause le caractère inéluctable de l'évolution.